SOUND BASED POSITIONING

THESIS

David L. Weathers, 2d Lt, USAF

AFIT-ENG-MS-17-M-081

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-17-M-081

SOUND BASED POSITIONING

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Electrical Engineering

David L. Weathers, B.S.E.E.

2d Lt, USAF

March 2017

AFIT-ENG-MS-17-M-081

SOUND BASED POSITIONING

THESIS

David L. Weathers, B.S.E.E.
2d Lt, USAF

Committee Membership:

Dr. John Raquet
Chair

Dr. Richard Martin
Member

Capt Aaron Canciani
Member

AFIT-ENG-MS-17-M-081

Abstract

With a growing interest in non-GPS positioning, navigation, and timing (PNT), sound based positioning provides a precise way to locate both sound sources and microphones through audible signals of opportunity (SoOPs). Exploiting SoOPs allows for passive location estimation. But, attributing each signal to a specific source location when signals are simultaneously emitting proves problematic. Using an array of microphones, unique SoOPs are identified and located through steered response beamforming. Sound source signals are then isolated through time-frequency masking to provide clear reference stations by which to estimate the location of a separate microphone through time difference of arrival measurements. Results are shown for real data.

Acknowledgements

First and foremost, thank you Jesus Christ for covering a debt I could never pay and for giving me new life in You. To my bride, I have loved our journey so far and I am truly humbled to have you by my side for the rest of our life - I love you! Mom and Dad, thank you for raising me to love learning, but more importantly, to love the One who holds all knowledge. My advisor, Dr. Raquet, thank you for your guidance and invaluable insight.

David L. Weathers

Table of Contents

Page

Abstract iv

Acknowledgements v

List of Figures ix

List of Tables xiii

I. Introduction 1

1.1 Alternative Positioning, Navigation, and Timing 1
1.2 Sound Based Positioning 1
1.3 Problem Statement 3
1.4 Thesis Overview 4

II. Background 5

2.1 Chapter Overview 5
2.2 Generalized Cross-Correlation for Time Difference of Arrival Measurements 5
2.3 Sound Source Arrangement Limitations 7
2.3.1 Dilution of Precision 8
2.3.2 Optimal Arrangement of Sound Sources and Microphones 8
2.4 Related Research 10
2.4.1 Section Overview 10
2.4.2 Cell Phone Location 10
2.4.3 Gunshot Detection 11
2.4.4 Human Speech Localization and Isolation 12
2.4.5 Location from Unmanned Aerial Vehicles 13
2.4.6 Acoustic Vector Sensors 14
2.4.7 Audio Array Calibration and Optimization 14

III. Positioning with Sequential Sound Sources at Known Locations 16

3.1 Chapter Overview 16
3.2 Tier 1 Methodology 17
3.2.1 Section Overview 17
3.2.2 Truth Data Collection 19
3.2.3 Signal Collection and Differentiation 20
3.2.4 Time of Arrival Estimation 21
3.2.5 Drift Correction 22

Page

3.2.6 Normalization of Data Through Correlated Additive White Gaussian Noise . . . 23
3.2.7 Microphone Position Estimation . . . 24
3.2.8 Methodology Summary . . . 27
3.3 Tier 1 Results . . . 28
3.3.1 Methods for Data Characterization . . . 28
3.3.2 Results by Testpoint . . . 31
3.4 Tier 2 Methodology . . . 41
3.4.1 Section Overview . . . 41
3.4.2 Time Difference of Arrival Estimation . . . 42
3.4.3 Location Estimation Through Closed Form Solutions . . . 42
3.4.4 Least Squares Location Estimation . . . 44
3.5 Tier 2 Results . . . 45
3.6 Chapter Summary . . . 52

IV. Positioning with Sequential Sound Sources at Unknown Locations . . . 54

4.1 Chapter overview . . . 54
4.2 Tier 3 Methodology . . . 55
4.2.1 Section Overview . . . 55
4.2.2 Location Estimation Through Cascading Closed Form Solution . . . 56
4.2.3 Least Squares Estimation for Speaker and Mobile Microphone Location . . . 57
4.3 Tier 3 Results . . . 59
4.3.1 Comparison of Dilution of Precision . . . 59
4.3.2 Tier 3 Results . . . 59
4.4 Tier 4 Methodology . . . 65
4.4.1 Section Overview . . . 65
4.4.2 Signal Collection and Differentiation . . . 65
4.4.3 Generalized Cross-Correlation Time Difference of Arrival Measurements . . . 66
4.5 Tier 4 Results . . . 67
4.6 Chapter Summary . . . 71

V. Positioning with Simultaneously Emitting Sound Sources at Unknown Locations . . . 72

5.1 Chapter Overview . . . 72
5.2 Tier 5 Methodology . . . 72
5.2.1 Section Overview . . . 72
5.2.2 Time Frequency Masking . . . 73

Page

5.2.3 Steered Response Power . . . 75
5.2.4 Peak Isolation Filtering . . . 77
5.2.5 Time Difference of Arrival Measurements . . . 80
5.2.6 Sound Source Selection . . . 81
5.3 Tier 5 Results . . . 82
5.4 Chapter Summary . . . 87

VI. Conclusion . . . 88

6.1 Research Summary . . . 88
6.2 Future Research and Applications . . . 90
6.2.1 Environmental Resiliency . . . 90
6.2.2 Sound Source Detection and Selection Methods . . . 90
6.2.3 UAV Detection through Steered Response Power Mapping . . . 91
6.2.4 Infrasound Positioning for Increased Scalability . . . 91
Bibliography . . . 92

List of Figures

Figure		Page
1	Received waveforms filtered, delayed, multiplied, and integrated for a variety of delays until peak output is obtained. Adapted from [17].	6
2	Illustration of DOP for range-based positioning [31]. (A) Two sound sources, with measured distances from the mic, creating finite solutions of the microphone location at intersections. (B) Same as A but showing errors on ranges, with the area of possible microphone locations shown in green. (C) Same as B but with poor DOP, creating a larger area of possible solutions.	9
3	Methodology for obtaining mobile microphone location estimates in Tier 1.	18
4	(a) Microphone with four reflective spheres tracked by Vicon cameras mounted on the wall in the background. (b) Virtual rendering of the microphone location and orientation in Vicon Software.	19
5	Example of single audio recording separated into five trials.	20
6	Peaks detected in temporally separated sound source playback	21
7	(a) Range measurements before drift correction due to receiver clock error. (b) Range measurements after drift correction is applied.	22
8	(a) Location estimation without correlated AWGN. The 100 estimates are constrained to six points due to quantization error. (b) Location estimation with correlated AWGN.	24
9	DOP map of testbed for Tier 1 and 2 tests. Color corresponds to DOP as a function of the location of the mobile microphone.	30
10	Locations of sound sources and testpoints for the mobile microphone	31

Figure Page

11 TOA measurement error distribution at Testpoint 0 32

12 Estimated location of mobile microphone at Testpoint 0 33

13 TOA measurement error distribution at Testpoint 1 34

14 Estimated location of mobile microphone at Testpoint 1 34

15 TOA measurement error distribution at Testpoint 2 35

16 Estimated location of mobile microphone at Testpoint 2 36

17 TOA measurement error distribution at Testpoint 3 37

18 Estimated location of mobile microphone at Testpoint 3 37

19 TOA measurement error distribution at Testpoint 4 38

20 Estimated location of mobile microphone at Testpoint 4 39

21 TOA measurement error distribution at Testpoint 5 40

22 Estimated location of mobile microphone at Testpoint 5 40

23 Methodology for obtaining mobile microphone location estimates in Tier 2. 41

24 Estimated location of mobile microphone at Testpoint 0 46

25 Estimated location of mobile microphone at Testpoint 1 47

26 Estimated location of mobile microphone at Testpoint 2 48

27 Estimated location of mobile microphone at Testpoint 3 49

28 Estimated location of mobile microphone at Testpoint 4 50

29 Estimated location of mobile microphone at Testpoint 5 51

30 Methodology for obtaining mobile microphone location estimates in Tier 3. 56

31 DOP map of testbed for Tier 3, 4, and 5 tests. Color corresponds to DOP as a function of the location of the mobile microphone. 60

32 Estimated location of mobile microphone at Testpoint 0 61

Figure Page

33 Estimated location of mobile microphone at Testpoint 1 62

34 Estimated location of mobile microphone at Testpoint 4 63

35 Estimated location of mobile microphone at Testpoint 5 64

36 Example of multi-channel audio recording for one trial. 66

37 Estimated location of mobile microphone at Testpoint 0 68

38 Estimated location of mobile microphone at Testpoint 1 69

39 Estimated location of mobile microphone at Testpoint 3 70

40 Methodology for obtaining mobile microphone location estimates in Tier 5. 73

41 TFM speaker of interest extraction system. Adapted from [30]. 75

42 Time averaged SRP map to showing all four sound source locations. 77

43 (a) Cross-section of unfiltered SRP map shown in Figure 42 at Y = 3 m. Estimated sound source locations (red) are both on a single peak. (b) SRP cross-section after PIF has been applied, producing correct sound source location estimates. 78

44 SRP map from Figure 42 with peak isolation filtering applied. Streaking has been reduced and four distinct peaks are shown corresponding to the source locations. 79

45 Modified GCC method, applying reconstructed audio for generating stronger, unambiguous correlation peaks. Peaks of the mobile microphone correlation and the reference microphone correlation are then differenced to determine the TDOA value. 80

46 (a) Cross-correlation example with good peak determination. (b) Cross-correlation example with ambiguous peak determination. 82

47 Estimated location of mobile microphone at Testpoint 0 83

48 Estimated location of mobile microphone at Testpoint 3 84

Figure Page

49 Estimated location of mobile microphone at Testpoint 4 85

50 Estimated location of mobile microphone at Testpoint 5 86

List of Tables

Table		Page
1	Progression of tiers towards more realistic scenarios.	4
2	Conditions of testing for Tiers 1 and 2.	16
3	Results for Test Performed in Tier 1	31
4	Results for Test Performed in Tier 2	45
5	Conditions of testing for Tiers 3 and 4.	55
6	Results for Test Performed in Tier 3	60
7	Results for Test Performed in Tier 4	67
8	Conditions of testing for Tier 5.	73
9	Results for Test Performed in Tier 5	82

SOUND BASED POSITIONING

I. Introduction

1.1 Alternative Positioning, Navigation, and Timing

In modern operations, the ability to perform precise Positioning, Navigation, and Timing (PNT) has become an indispensable asset. While the reliance on the Global Positioning System (GPS) continually increases in order to meet the demands of PNT-critical operations, so does interest in alternative PNT methods. In circumstances where GPS capabilities are unavailable or degraded beyond operation requirements, alternative PNT may provide the necessary capabilities. Examples of alternative PNT include vision [24], magnetic field anomalies [10], received signal strength from cell towers [15], and sound [22]. A benefit of many of these methods is the ability to perform passive navigation. While GPS requires signal transmission from satellites, many alternative PNT methods avoid broadcasting signals by exploiting existing signals or features readily available in their respective environments. While Signals of Opportunity (SoOPs) are typically understood as radio frequency signals already available in the environment that can be used for navigation, SoOPs are not limited to radio frequencies.

1.2 Sound Based Positioning

With the growing interest in alternative PNT, sound based positioning provides a precise way to locate both sound sources and sensors through audible SoOPs. The

abundance of SoOPs in the audible frequency range provides ample information about the environment. The unconscious familiarity of sound based positioning to everyday life allows most people to seamlessly glean understanding about their environment through stereophonic hearing. Even with closed eyes, one can listen to determine the general origin of a sound source. The brain determines the location of the sound source based on the minute difference in time taken for the sound to propagate to each ear [6]. In fact, in most modern musical recordings, the stereo tracks of each instrument are slightly offset in order to replicate to the listener how the music would sound if heard live. Just as human hearing allows sound source location detection, many computer systems perform the same task through arrays of microphones, but with much greater accuracy and precision. Solutions using microphone arrays are not limited to two sensors the way humans are with two ears; computers can utilize hundreds of microphones over a given field and calculate the time delays of every microphone [23]. Also, microphones can be spaced much further apart than the width of the head, which humans are limited to with the spacing of ears. This allows for larger areas of observation. Computer-based systems have proven beneficial in sound-based positioning applications such as locating sound sources over a large area [2], simultaneous localization of several sound sources [4], increasing the intelligibility of sound sources [30], and determining the location of enemy fire in combat [9]. While the possibility of sound based navigation replacing the capabilities of GPS is not a thought on the horizon, there are several circumstances where it may perform more accurately, or produce residual effects, such as audio surveillance, that GPS cannot match.

1.3 Problem Statement

The goal of this research was to explore the capabilities of sound based positioning as an alternative form of PNT. As a means to exploring location estimation via sound, a system was developed capable of locating a mobile microphone by using an array of reference microphones to capture audio from sources of unknown location. Because passive location is one of the main advantages of GPS-alternative PNT, the system was designed to perform without generating its own sound by exploiting SoOPs common to the audible range. Because of the multiple complexities involved in completing the task, the design process was broken into five tiers. Each tier progression holds fewer assumptions and thus becomes more applicable to real-world scenarios, with Tier 5 meeting the original design requirements. The conditions of testing are described in Table 1. Sound source type indicates the waveform that each of the sound sources generates. Sound source timing indicates whether or not the solution assumes the sound sources are transmitted at exactly known intervals. Sound source location indicates if a priori knowledge of the source locations are available for use in estimating the location of the mobile microphone. Successive playback assumes each sound source completes its transmission before another sound source begins playing, whereas simultaneous playback assumes multiple sound sources may be overlapping in their transmission.

Table 1. Progression of tiers towards more realistic scenarios.

Tier	Sound Source Type	Sound Source Timing	Sound Source Location	Playback
1	Impulse	Known	Known	Successive
2	Impulse	Unknown	Known	Successive
3	Impulse	Unknown	Unknown	Successive
4	Recorded Speech	Unknown	Unknown	Successive
5	Recorded Speech	Unknown	Unknown	Simultaneous

1.4 Thesis Overview

This document is composed of six chapters. Chapter II presents pertinent mathematical background to the methods used in testing as well as a review of related research. Chapter III presents both the methodology and results of Tiers 1 and 2, where the locations of the sound sources are known and emit sequentially. Chapter IV presents the methodology of and results of Tiers 3 and 4, which allow for positioning using sound sources of unknown position that emit sequentially. Chapter V presents the methodology and results of Tier 5, which allows for positioning using sound sources of unknown location and simultaneous emission. Lastly, Chapter VI provides a conclusion of research and discusses the potential for future related work.

II. Background

2.1 Chapter Overview

This chapter provides a background on the differentiation and estimation methods employed, optimal speaker placements for sound based positioning, and recent research related to sound based positioning. Section 2.2 covers Generalized Cross-Correlation (GCC) Time Difference of Arrival (TDOA) estimation methods. Section 2.3 covers how the geometric arrangement of sound sources and microphones affects the accuracy of location estimation and how the arrangement may be optimized for best results. Finally, Section 2.4 reviews research related to sound based positioning.

2.2 Generalized Cross-Correlation for Time Difference of Arrival Measurements

Knapp and Carter proposed the generalized correlation method for estimation of time delay in 1976, which has been the pivotal reference for time delay estimation between spatially separated sensors [5, 17]. A maximum likelihood estimator is developed as a pair of receiver prefilters for two signals followed by a cross correlator. These signals are modeled as

$$a_i(t) = s_i(t) + n_i(t), \qquad a_j(t) = s_j(t) + n_j(t), \tag{1}$$

where $s(t)$ represents a real signal and $n(t)$ represents uncorrelated noise [17]. While the GCC method assumes stationarity of both $s(t)$ and $n(t)$, it is commonly employed in slowly varying environments where the signal and noise remain stationary over the finite observation time [17]. This method for determining the time delay between signals seeks the maximal value of the cross correlation function:

$$\hat{\rho} = \arg\max_{\tau} R_{a_i a_j}(\tau) = \arg\max_{\tau} \mathcal{E}\left\{[a_i(t)a_j(t-\tau)]\right\} \tag{2}$$

where $\hat{\rho}$ is the estimate of the true time delay ρ and $\mathcal{E}\{\}$ denotes expectation [17]. However, with only a finite observation time, $R_{a_i a_j}(\tau)$ must be estimated:

$$\hat{R}_{a_i a_j}(\tau) = \frac{1}{T-\tau}\int_{\tau}^{T} a_i(t)a_j(t-\tau)dt, \tag{3}$$

where T is the observation interval [17]. In order to produce a more accurate estimate when a priori knowledge of the signal and noise statistics are available, several filtering methods, including the Hannan-Thompson processor [17] may be applied to $a_i(t)$ and $a_j(t)$. These methods filter the signals in order to weight the cross correlation computation to frequencies with a higher Signal to Noise Ratio (SNR) in order to give a less varied estimate of the time delay between sensors. The filters are shown in Figure 1 as H_i and H_j, which modify the received waveforms, a_i and a_j, into b_i and b_j respectively. b_j is then delayed and multiplied with b_i, integrated, and squared for a range of time shifts τ until a peak, $\hat{\rho}$ is discovered.

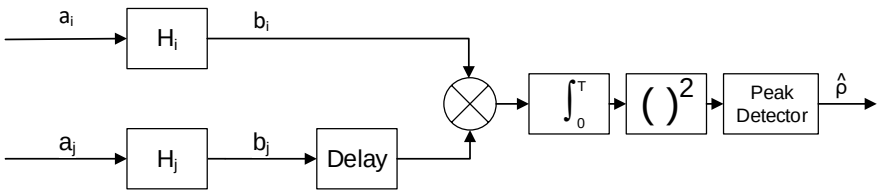


Figure 1. Received waveforms filtered, delayed, multiplied, and integrated for a variety of delays until peak output is obtained. Adapted from [17].

When the filters H_i and H_j are uniform across the sampled spectrum, $\hat{\rho}$ is the max of the cross-correlation of a_i and a_j as shown in Equation (2). The Hannan-Thompson filtering method sets the weight of the filters $\psi \triangleq H_i H_j^*$ such that

$$\psi = \frac{|\gamma_{ij}|^2}{\left|G_{a_i a_j}\right| \left[1 - |\gamma_{ij}|^2\right]}, \tag{4}$$

where G is the cross-power spectral density of the subscripted signals and γ_{ij} is the coherence estimate given as [17]

$$\gamma_{ij} \triangleq \frac{G_{a_i a_j}}{\sqrt{G_{a_i a_i} G_{a_j a_j}}}. \tag{5}$$

Filters such as the Hannan-Thompson filter help eliminate ambiguities caused by narrowband sounds when microphone spacing is too far apart [2]. Under ideal conditions, the Hannan-Thompson processor achieves the Cramér-Rao lower bound on variance for delay estimators. However, as the SNR decreases, so does the effectiveness of filtering techniques for delay estimation [17]. With a priori knowledge on the spectra of the signal, filtering processes may allow for significantly improved time delay estimates by obtaining non-integer delays from the sampling frequency [18]. Unless some characteristic about the spectra of the sound source is available, the proposed filtering methods offer little to no advantage compared to the generalized cross correlation estimate proposed in Equation (2).

2.3 Sound Source Arrangement Limitations

The ability to determine an accurate location estimate is limited by the geometric arrangement of the microphones. As more indoor localization solutions are going beyond proof of concept and prototype states, localization error should be given more consideration considering sensor placement is often done by hand, leading to otherwise avoidable inaccuracies [16].

2.3.1 Dilution of Precision.

The uncertainties stemming from the geometric arrangement of the microphones can be quantified through Dilution of Precision (DOP) values. DOP indicates how much the fundamental ranging error is magnified by the geometric relation among the speaker and microphone positions. Assuming all measurements of $\hat{\boldsymbol{\rho}}$ have the same variance and those measurements are uncorrelated, DOP may be derived as

$$DOP = \sqrt{\mathrm{Tr}\left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right]}, \tag{6}$$

where Tr indicates the trace matrix operation and $\mathbf{H}$ is a matrix containing the derivatives of the range difference equations with respect to the unknown state vector. $\mathbf{H}$ is more explicitly defined in Equation (18) of Section 3.2.7. Smaller DOP values correspond to areas where microphone positioning estimation may be more accurate, as shown in Figure 2B. Figure 2C depicts large DOP values corresponding to less accurate microphone positioning estimations.

Since the proposed solution is limited to estimating a microphone on a 2D plane and the related timing errors, Horizontal and Time Dilution of Precision (HTDOP), a more specific form of DOP is considered. HTDOP only accounts for variance in 2D coordinates and clock error of the $\mathbf{H}$ matrix.

2.3.2 Optimal Arrangement of Sound Sources and Microphones.

Concerning the geometric shape of a four element array, [25] shows that in an ideal environment with minimized reverberation, a square array configuration of microphones provides optimal estimation accuracy in locating a sound source within the area of the array. However, as reverberation increases, the ability to accurately estimate the location of the sound source is dramatically affected. Thus, in noisy, especially reverberant, environments, irregularly shaped array geometries often out-

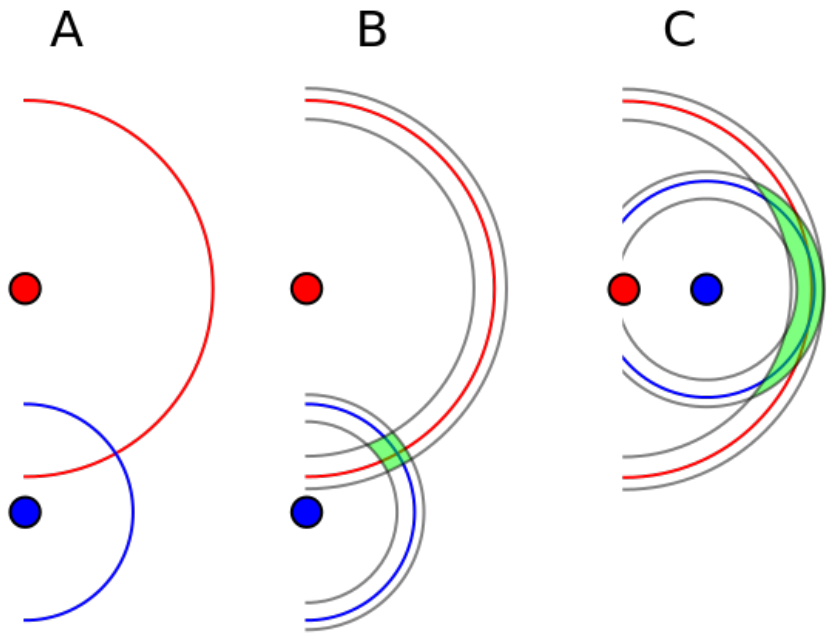


Figure 2. Illustration of DOP for range-based positioning [31]. (A) Two sound sources, with measured distances from the mic, creating finite solutions of the microphone location at intersections. (B) Same as A but showing errors on ranges, with the area of possible microphone locations shown in green. (C) Same as B but with poor DOP, creating a larger area of possible solutions.

perform regularly shaped arrays [14].

Another aspect to consider in designing the microphone and sound source array shape is the distance between sound sources. There is a trade-off between relatively small spaced arrays, which reduce ambiguity due to sound source wavelengths being shorter than the spacing of the microphones, and large spaced arrays, which increase resolution [3].

The optimal microphone array is highly dependent on the environment in which the system is implemented, so there is not a single best solution for all real-world applications. Because of this, heuristic methods such as the Genetic Algorithm are implemented in order to select a near-optimal placement of a set number of microphones and sound sources in a given room [14].

2.4 Related Research

2.4.1 Section Overview.

There are several areas of research within sound based navigation. Due to the ubiquity of cell phones with hard-wired speakers and microphones, the ability to locate a cell phone through sound based methods is of interest. Many signals of interest have concentrated power within the audible range, including gunshots and human speech, which allows for a sound based navigation system to passively detect these signals. Sound based methods have been tested on board Unmanned Aerial Vehicles (UAV)s for collision avoidance and emergency vehicle detection. There are also new sensor technologies being developed, including the Acoustic Vector Sensor (AVS), as well as array calibration methods. Section 2.4.2 discusses locating cell phones via sound and other signals of opportunity. Section 2.4.3 discusses research in gunshot detection. Section 2.4.4 discusses research in detecting and locating human speech. Section 2.4.5 discusses developments in sound-based positioning from UAVs. Section 2.4.6 presents resent research involving sound location via the AVS, and finally, Section 2.4.7 discusses microphone and speaker array calibration and optimization.

2.4.2 Cell Phone Location.

With the unavailability of GPS in indoor scenarios, alternative methods for cell phone navigation navigation have been a focus of research. By combining data from several sensors typically available of cell phones, including accelerometers, magnetic field sensors, gyroscopes, and sound sensors, [12] proposes a method to position a cell phone through dead reckoning until GPS is available again. During testing, location estimates showed on average a 1.8 m deviation from the true cell phone location after traveling 50 m.

In [26], Schuller et al. present a sound based navigation technique that tracks

cellphone-equipped bicyclists on their route. Instead of estimating the exact coordinates of the bicyclist, the solution only provides categorical results - a general section of the bike path. The audio from the cyclist's cellphone as he passes through a given portion of the route is compared with typical auditory scenery of each section of the route.

In [11], audio beacons are set up to allow cell phones to determine their location in an indoor environment. Similar to [26], instead of estimating exact coordinates, the results only estimate the room in which the cell phone is, not the location within the room. In addition, the audio beacons must contain a recognizable code - not just ambient noise or people speaking within the room. Therefore, the approach only allows for a cooperative relationship between the sound source and the cell phone. However, the novelty of the research lies in the cost-effectiveness and wide adaptability to mobile devices, including older and low-end models.

2.4.3 Gunshot Detection.

By comparing the TDOA of the audio signals from spatially separated microphones, the location of the gunshot can be determined. Though a well-researched field [1, 9, 22, 32, 33], gunshot detection is continually improving through sound-based techniques that not only locate the source of the gunshot, but also classify what type of weapon was fired. [32] has improved the accuracy of gunshot location estimation and classification by as much as 30% compared to current methods by modeling the natural noise of the environment as Symmetric Alpha Stable (SAS) instead of the usually assumed Gaussian noise according to the Ensemble Empirical Mode Decomposition data analysis method. The SAS distribution can be very similar to the Gaussian distribution, but better models heavy tailed phenomena, such as noise in the audible frequency range [7, 32].

2.4.4 Human Speech Localization and Isolation.

Of particular interest to the field of sound-based navigation research is tracking human speech. The two roots of the problem are determining the location of the human subject, and once the subjects have been located, increasing the intelligibility of speech for the speaker of interest. One method for determining the location of sound sources is Steered Response Power (SRP) mapping [13]. Within a given area of interest, the audio from an array of spatially distributed microphones is delay and sum beamformed. Beamforming delays the audio from spatially separated microphones according to the TDOA measurements so that all audio channels constructively combine to raise the signal power at areas of coherent sound sources [20]. After the locations of sound sources are determined, [30] implements Time-Frequency Masking (TFM) techniques on the signal of interest, which raises the signal to noise ratio by maintaining areas of the time frequency map corresponding to the signal of interest and eliminating areas corresponding to noise sources. Results are measured according to a Speech Intelligibility Index (SII), where a SII of < 0.25 indicates unintelligible speech, $0.25 <$ SII < 0.5 indicates speech intelligibility with concentrated listening, and SII > 0.5 indicates clear speech intelligibility with eased listening [30]. The results show that determining the locations of noise sources and masking the interference those sources create significantly improves the SII compared to only beamforming without masking, and in many cases increased the SII past the critical threshold of 0.25. The primary disadvantage is the computational burden of simultaneously beamforming on multiple sound sources, which currently prevents real-time applications. However, the approach is well-suited for localization of the signal of interest and enhancement of the SII in applications where post processing is acceptable. Both SRP mapping and TFM are implemented and explained in more detail in Chapter V.

2.4.5 Location from Unmanned Aerial Vehicles.

A critical aspect of employing a flight of UAV is ensuring collision avoidance through formation control. A low computational cost solution is presented in [3] to provide each UAV with the relative position of surrounding UAVs. Each UAV was equipped with a four channel array of microphones in order to determine the location of the source generated by piezoelectric transducers on other cooperative UAVs. Future work proposes detecting the sound from the engine of other UAVs in order to expand capabilities to locating non-cooperative UAVs.

In addition to locating other aerial vehicles, sound based navigation techniques are being applied to locate narrow-band signals on the ground with hopes of locating emergency distress signals from whistles attached to most aircraft life vests [2]. This technique would allow a search team UAV to locate their subjects in night time, through foliage or dust, fog, and smoke. The test used the same UAV and attached microphone array from [3], but implemented a particle filtering localization technique to recursively estimate the target location. The proposed technique overcame ambiguities in the TDOA measurements through knowledge of the UAV trajectory and by computing the relative velocity from Doppler shift of the known distress signal. At a range of 150 meters, the UAV successfully located both an emergency safety whistle and a piezo alarm, both emitting constant, narrowband frequencies.

In [21], a UAV is equipped with a 16 channel microphone array and location estimations are calculated through MUltiple SIgnal Classification (MUSIC). MUSIC allows for accurate localization of sound sources by whitening high power noise through implementing a noise correlation matrix. However, MUSIC as originally designed can only account for spatially static noise sources. Several improvements to MUSIC are compared according to likelihood of detection and the computational burden each method imposed. One of the proposed improvements, MUSIC based on incremental

Generalized Singular Value Decomposition (iGSVD-MUSIC), whitens noise similarly to MUSIC, but allows for dynamic tracking of the noise sources and drastically reduces the computational cost without sacrificing sound source localization accuracy. While all results were post-processed, the lower computational cost of iGSVD-MUSIC would allow for future research involving real-time sound source detection from a UAV.

2.4.6 Acoustic Vector Sensors.

The recent technological advancement of the AVS, developed by Microflown Technologies, requires far fewer sensors in an array compared to a traditional microphone array. Typical microphones only measure sound pressure, the scalar component of sound waves. The AVS also measures the the vector component of sound waves, acoustic particle velocity. As the sound wave passes two parallel, thin, heated wires, the first wire cools before the second, giving an indication of the acoustic particle velocity in one dimension. With three orthogonal pairs of wires, the AVS provides a 3d velocity. [4]. By measuring both sound pressure and acoustic particle velocity, a single AVS can determine the direction of a sound source. Given a moving sound source, one AVS can estimate the distance to the sound source as well [4]. With an array of AVSs, the accuracy of sound source location estimates are comparable to traditional microphone arrays with many more channels [8]. Several applications of the AVS have been researched, including the localization and tracking of aircraft [8], multiple sound source tracking [4], RPG detection, and sense and avoid capabilities onboard UAVs [9].

2.4.7 Audio Array Calibration and Optimization.

Because sensor placement by hand often leads to measurement inaccuracies [16], an approach is presented in [23] to directly recover the location of both microphones

and sound sources in a 3D environment from TDOA measurements. By locating all of the microphones and sound sources, the system can be calibrated relative to the location of one of the sensors, limiting sensor placement inaccuracy. The proposed method requires an array of at least 10 microphones and 5 sound sources or 10 sound sources and 5 microphones in order to simultaneously solve for the locations of all components in the array. When the solution is restricted to a 2D environment, at least 8 microphones and 4 sound sources or 8 sound sources and 4 microphones are required. The approach has been tested using simulated data with no noise and recovers positions of the microphones and speakers relative to a reference microphone. The next stage of research consists of expanding testing to real world data.

While the accuracy of location estimates in sound-based navigation is crucial, the processing time to compute the estimates must also be considered in order to implement real-time solutions. If the estimate takes too long to calculate, the estimate may not provide any useful information on the location of the object of interest. As the number of sensors in the array increases, the computational burden of processing a solution increases exponentially. In order to reduce the time to estimation, [16] developed a method to determine a sensor arrangement with a minimal number of sensors for a given sound source configuration that provides an estimate within 1.5 times the optimal solution.

III. Positioning with Sequential Sound Sources at Known Locations

3.1 Chapter Overview

As the initial stage, the results of Tier 1 provide a proof of concept for sound based navigation and set the foundation for more complicated real-world scenarios. The goal of Tier 1 was to locate the mobile microphone based on the differences in arrival time of the signals from four sound sources. In this tier, the sound sources were speakers with known locations, emitting a short duration impulse as the signal. The four sound sources played the signal successively at precisely timed intervals, which eliminated the need for signal source differentiation. The order in which the signals played determines the sound source from which sound source the signal originated. The tests were performed indoors in a low-noise facility using a high fidelity MXL-604 microphone as the object to be located.

In Tier 2, the time that the sound sources emit signals relative to one another was unknown. In order to maintain accurate positioning capabilities with unknown signal timing, the solution for Tier 2 introduced a reference microphone with known location. Comparing the audio of the reference and mobile microphone allows for TDOA calculations as the basis for the location estimation instead of Time of Arrival (TOA) measurements. Table 2 summarizes the conditions of testing for Tiers 1 and 2.

Table 2. Conditions of testing for Tiers 1 and 2.

Tier	Sound Source Type	Sound Source Timing	Sound Source Location	Playback
1	Impulse	Known	Known	Successive
2	Impulse	Unknown	Known	Successive

This chapter covers the methodology used in Tier 1 testing and data processing in Section 3.2, followed by the results and analysis of data from Tier1 in Section 3.3. Section 3.4 discusses the methodology improvements to allow for TDOA-based location estimation. Section 3.5 presents the results of Tier 2 tests and Section 3.6 summarizes the findings of Tiers 1 and 2.

3.2 Tier 1 Methodology

3.2.1 Section Overview.

This section covers covers the methodology of data collection and location estimation for Tier 1. Many of the methods presented in this section are also implemented in later tiers. Topics discussed in this section include truth data collection, audio signal collection and differentiation of trials and sound sources, TOA measurement formation, receiver clock drift correction, normalization of TOA measurements, and estimation of the mobile microphone position from TOA measurements. Figure 3 visualizes both the test area and the progression of the solution for Tier 1 tests.

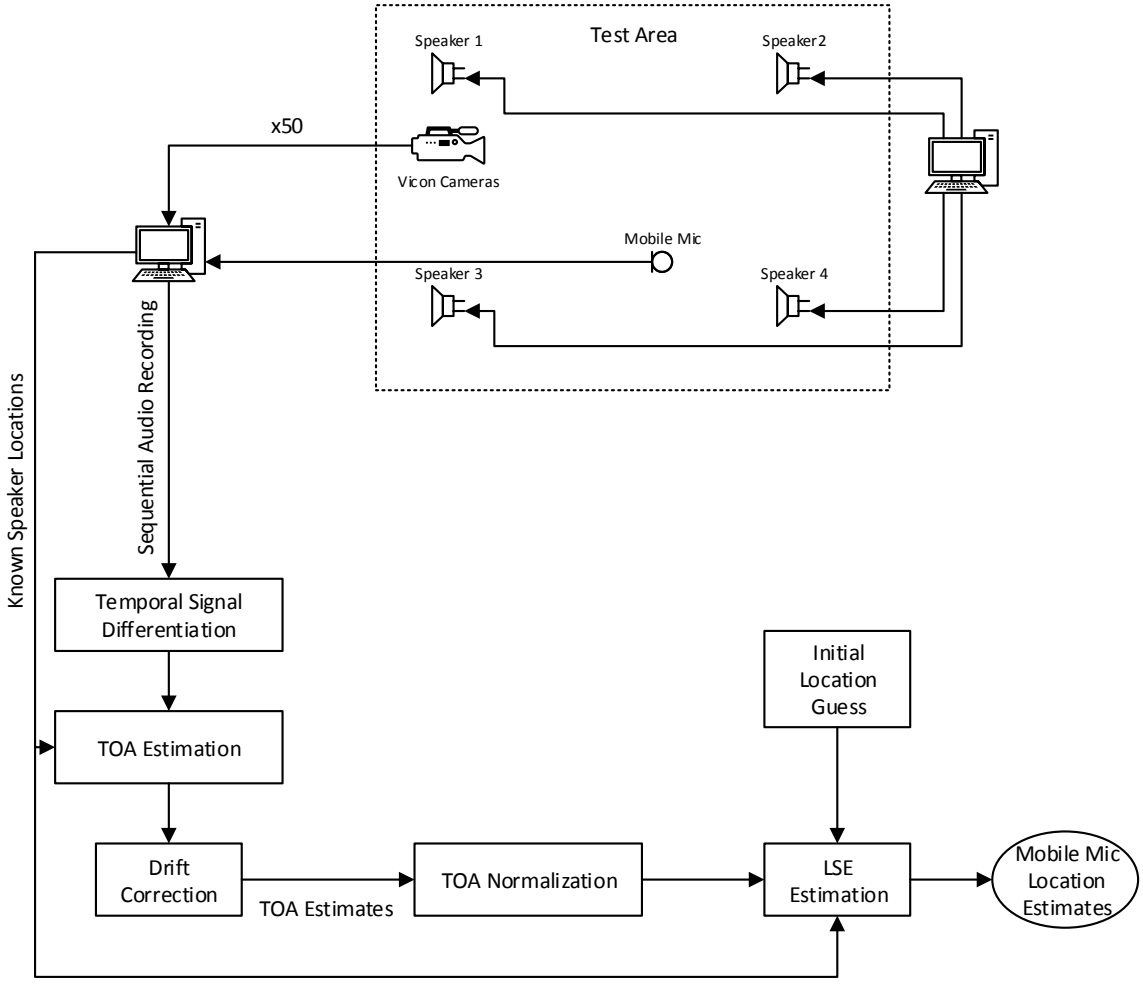
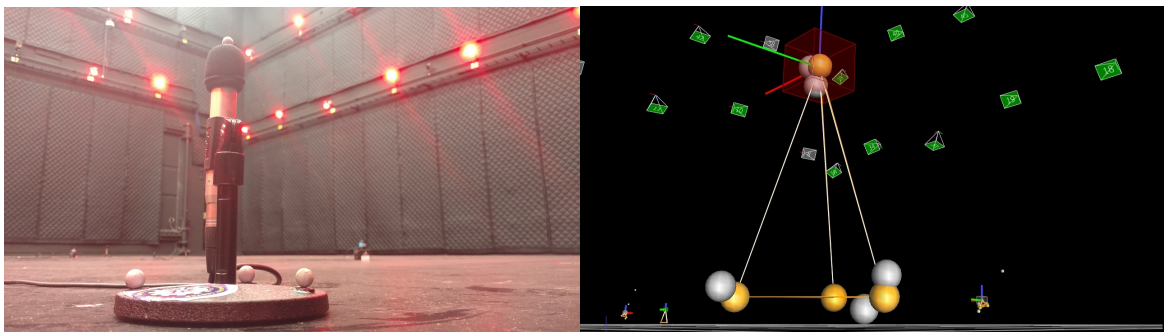


Figure 3. Methodology for obtaining mobile microphone location estimates in Tier 1.

3.2.2 Truth Data Collection.

A Vicon motion capture system provided truth data on the positions of the four sound sources and the mobile microphone. Each object was affixed with four reflective sphere markers, shown in Figure 4a, which the Vicon cameras tracked to determine the location and orientation of the object. One of the four reflective markers is located at the center of the speaker cone or microphone head, which was designated as the reference coordinates of the object, shown by the intersection of the colored axes in Figure 4b. Prior to sound based testing, the Vicon system produced one hundred location samples of each stationary object. The averages of the samples are assumed to be the true known coordinates of each object. Using the true coordinates, the true distance between each object was calculated as a standard for the estimated distances.



(a) **(b)**

Figure 4. (a) Microphone with four reflective spheres tracked by Vicon cameras mounted on the wall in the background. (b) Virtual rendering of the microphone location and orientation in Vicon Software.

3.2.3 Signal Collection and Differentiation.

An impulse signal was played through each speaker successively; after the signal played from the first speaker, the remaining three speakers played the same signal 0.5 seconds after the previous speaker. Successive playback of the signals eliminated the need for more robust signal differentiation methods by temporally separating the signals. A more complex solution that successfully differentiates between simultaneously emitting sound sources is proposed in Chapter V. After all four signals have played, two seconds of silence separated the end of the trial from the beginning of the next. 100 trials were performed for each testpoint. Once all trials of a testpoint were complete, the recording was then sectioned into the four second-long trials as shown in Figure 5.

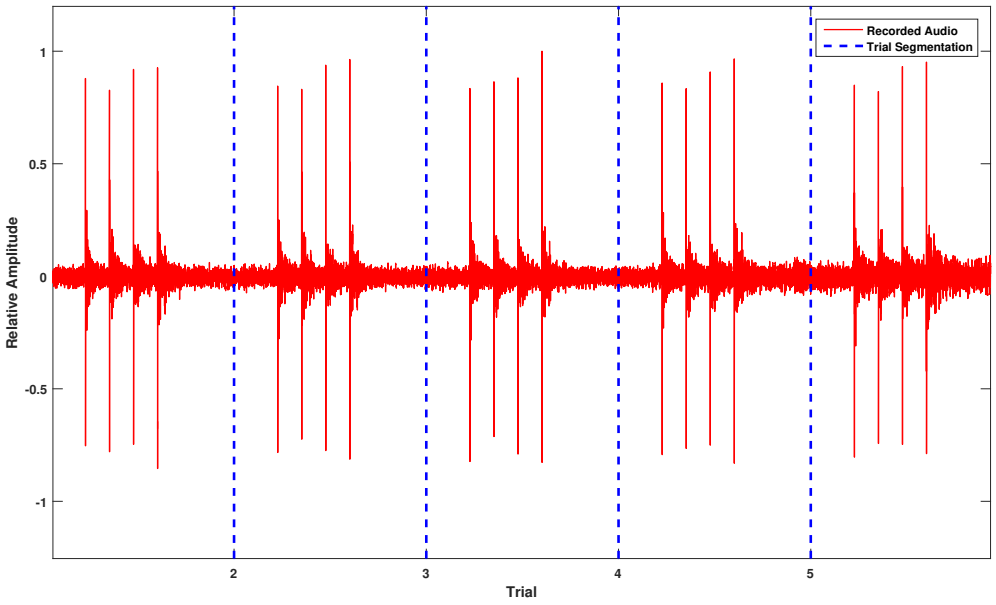


Figure 5. Example of single audio recording separated into five trials.

3.2.4 Time of Arrival Estimation.

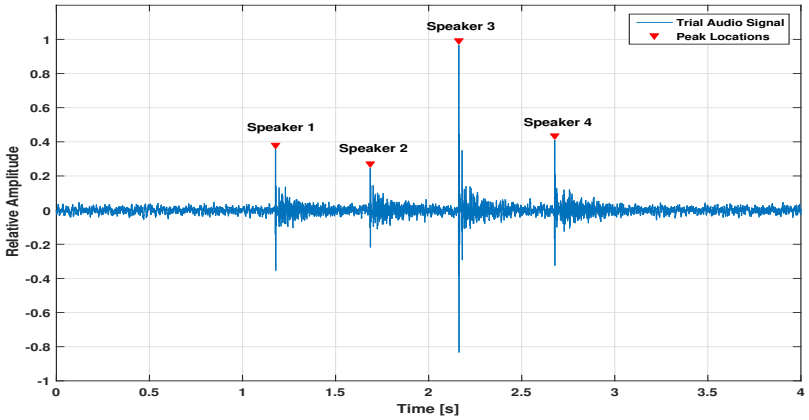


Figure 6. Peaks detected in temporally separated sound source playback

In each trial, the times corresponding to the four maximum values spaced at least 0.5 seconds apart were set to be $\mathbf{T}_a$, the measured arrival time of the signals from the sound sources to the mobile microphone. Example values in the vector $\mathbf{T}_a$ are shown in Figure 6 as the peak location markers. Arrival times are measured relative to a constant but arbitrary transmission time for each sound source, $\mathbf{T}_t$. Multiplying by the speed of sound, c, forms the range estimates of the trial:

$$\boldsymbol{\rho} = c\ (\mathbf{T}_a - \mathbf{T}_t). \tag{7}$$

Because the TOA measurement is dependent on the speed of sound, inaccuracies in the estimating the speed of sound affect the TOA estimate. Before each test was performed, the temperature was measured in order to calculate the speed of sound in the the test facility:

$$c = 331.3\sqrt{1 + \frac{\theta_c}{273.15}}\ \ [\text{m/s}] \tag{8}$$

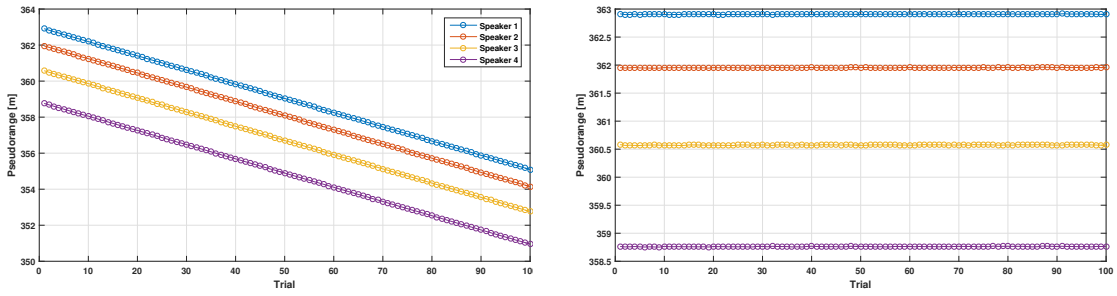
where θ_c is temperature in degrees Celsius [27].

3.2.5 Drift Correction.

In TOA based solutions as used in Tier 1, the accuracy of the results depends on the consistency of the receiver clock so that the recorder captures audio exactly at the sampling frequency, 44100 Hz for the duration of the recording. Because the signals from each of the four speakers played sequentially in precise intervals, receiver clock error caused TOA values to appear as if the signal was emitted slightly before or slightly after the expected interval. The receiver clock had a near linear drift, approximated by the mean difference in arrival times between each trial and the following trial:

$$drift \approx \frac{1}{N-1} \sum_{n=1}^{N-1} \frac{\boldsymbol{\rho}_n - \boldsymbol{\rho}_{n+1}}{\Delta t} = \frac{\boldsymbol{\rho}_1 - \boldsymbol{\rho}_N}{(N-1)\Delta t}, \tag{9}$$

where n represents the trial number, N is the total number of trials and Δt is the duration of each trial. This drift value was used to correct TOA estimates for both drift between trials as well as drift between signals from the four sound sources within the same trial. The effect of the drift corrections are shown in the comparison of Figures 7a and 7b. Because subsequent tiers all used TDOA-based solutions where clock errors cancel out in differencing, drift correction was only necessary in Tier 1.

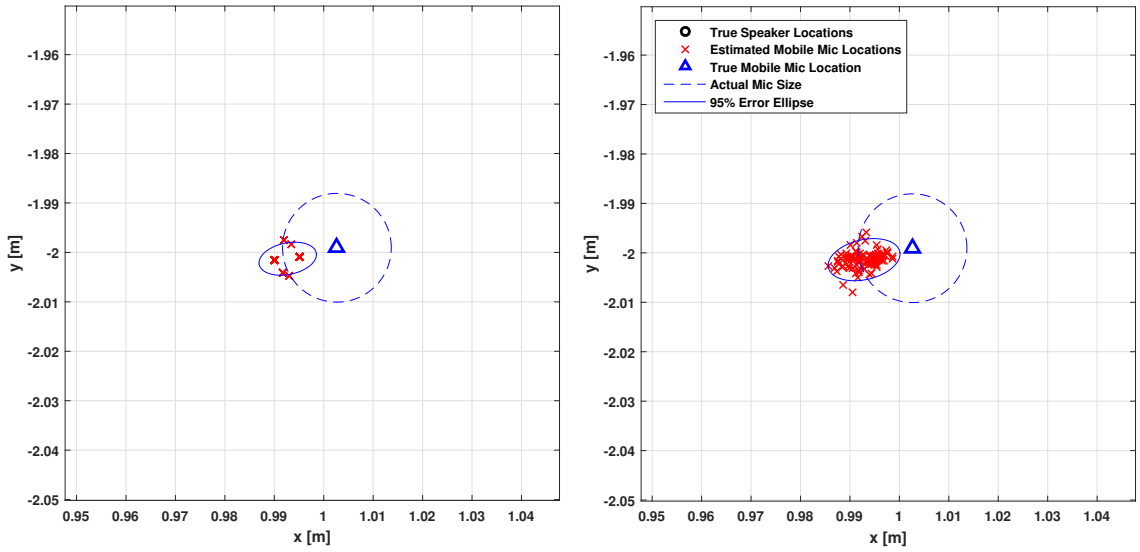


(a) **(b)**

Figure 7. (a) Range measurements before drift correction due to receiver clock error. (b) Range measurements after drift correction is applied.

3.2.6 Normalization of Data Through Correlated Additive White Gaussian Noise.

Assuming no quantization error, the variance in range measurements was primarily caused by thermal and environmental noise. Given the assumption is true, the distribution of measurements may reasonably be assumed to follow a zero mean Gaussian model [29]. However, the signals were recorded in the audible frequency range, with a sampling rate of $f_s = 44.1$ kHz. Therefore, the resolution of range measurements is limited to $\pm \frac{c}{2fs}$, approximately ± 4 mm. In an otherwise quiet room, the variation of measurements was often smaller than the resolution due to the sampling rate, causing quantization error. With quantization error, the distribution of location estimates concentrated to sparse discrete values. In order to make the distribution follow a more Gaussian pattern, zero mean Additive White Gaussian Noise (AWGN) with a covariance equal to the covariance of the original estimates was added to estimates. The AWGN must be properly correlated in order to maintain the orientation and eccentricity of the error ellipse, as shown in Figure 8. Adding noise also increased the variance in location estimates, causing the error ellipse to be slightly larger. While the increased variance in the location estimates is not ideal, AWGN allowed use of DOP based analysis as described in Section 3.3.1.



(a) (b)

Figure 8. (a) Location estimation without correlated AWGN. The 100 estimates are constrained to six points due to quantization error. (b) Location estimation with correlated AWGN.

3.2.7 Microphone Position Estimation.

The normalized and drift corrected range measurements from the k^{th} sound source to the mobile microphone, m, are distributed as

$$\boldsymbol{\rho}_m^s = [\rho_m^{s_1}, \rho_m^{s_2}, \cdots, \rho_m^{s_k}, \cdots, \rho_m^{s_K}]^T, \tag{10}$$

where K represents the total number of sound sources. The location of the mobile microphone, $[x_m\ ,\ y_m]^T$, is always unknown. Including receiver clock error, the unknown state vector contains all quantities to be estimated:

$$\boldsymbol{x} = \begin{bmatrix} x_m \\ y_m \\ \delta t \end{bmatrix}, \tag{11}$$

where δt is the clock error measured as a distance, relative to the speed of sound. In Tier 1, the locations of the K sound sources,

$$\begin{bmatrix} \boldsymbol{x}_s & \mathbf{y}_s \end{bmatrix} = \begin{bmatrix} x_{s_1} & y_{s_1} \\ x_{s_2} & y_{s_2} \\ \vdots & \vdots \\ x_{s_K} & y_{s_K} \end{bmatrix}, \tag{12}$$

are known values according to the truth data acquired through the Vicon system. Range measurements from the k^{th} sound source to the mobile microphone are related to $\boldsymbol{x}$ and $\boldsymbol{x}_s$ as

$$\rho_m^{s_k} = \sqrt{(x_{s_k} - x_m)^2 + (y_{s_k} - y_m)^2} + \delta t + \upsilon_k \tag{13}$$

where υ_k represents white Gaussian noise associated with the k^{th} sound source. Equation (13) can be rewritten as

$$\rho_m^{s_k} = h(\boldsymbol{x}) + \upsilon_k. \tag{14}$$

$$h(\boldsymbol{x}) = \sqrt{(x_{s_k} - x_m)^2 + (y_{s_k} - y_m)^2} + \delta t. \tag{15}$$

Represented in matrix form to account for range measurements from all measured sound sources,

$$\boldsymbol{\rho}_m^s = \mathbf{h}(\boldsymbol{x}) + \boldsymbol{\upsilon}. \tag{16}$$

$\mathbf{h}(\boldsymbol{x})$ is a nonlinear vector function [29] which may be solved through both closed form or estimated solution techniques. In Tier 1, an iterative Least Squares Estimation

(LSE) method is applied.

In order to determine a simple approximate solution to estimate $\boldsymbol{x}$, $\mathbf{h}(\boldsymbol{x})$ can be linearized by considering only the first two terms of the Taylor series expansion about a reference point,

$$\boldsymbol{x}_0 = \begin{bmatrix} x_0 & y_0 & \delta t_0 \end{bmatrix}, \tag{17}$$

such that $\mathbf{H}$ is a matrix of partial derivatives of the measurement vector with respect to the state vector, evaluated at $\boldsymbol{x}_0$:

$$\mathbf{H} = \left.\frac{\partial \mathbf{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}. \tag{18}$$

Note that $\boldsymbol{x}_0$ should be close enough to the real value of $\boldsymbol{x}$ such that the linearization of h is a valid assumption [29]. In Tier 1, $\boldsymbol{x}_0$ is initialized at the center of the test area. However, in the remaining tiers, where TDOA measurements are used, a closed form solution provides a more accurate initial $\boldsymbol{x}_0$. In the case of Tier 1, $\mathbf{H}$ may be explicitly defined as:

$$\mathbf{H} = \begin{bmatrix} \left.\frac{\partial h_1(\boldsymbol{x})}{\partial x_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & \left.\frac{\partial h_1(\boldsymbol{x})}{\partial y_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & \left.\frac{\partial h_1(\boldsymbol{x})}{\partial \delta t}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} \\ \vdots & \vdots & \vdots \\ \left.\frac{\partial h_K(\boldsymbol{x})}{\partial x_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & \left.\frac{\partial h_K(\boldsymbol{x})}{\partial y_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & \left.\frac{\partial h_K(\boldsymbol{x})}{\partial \delta t}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} \end{bmatrix}, \tag{19}$$

where the subscript on h denotes which sound source the derivative is taken relative to. Since δt is linear with respect to the reference point, $\left.\frac{\partial h(\boldsymbol{x},k)}{\partial \delta t}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}$ is 1 for all k. Equation (19) can be simplified to

$$\mathbf{H} = \begin{bmatrix} \left.\frac{\partial h_1(\boldsymbol{x})}{\partial x_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & \left.\frac{\partial h_1(\boldsymbol{x})}{\partial y_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & 1 \\ \vdots & \vdots & \vdots \\ \left.\frac{\partial h_K(\boldsymbol{x})}{\partial x_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & \left.\frac{\partial h_K(\boldsymbol{x})}{\partial y_m}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} & 1 \end{bmatrix}. \tag{20}$$

We define the error in the nominal estimation values, $\Delta\boldsymbol{x}$, as used in the least squares solution solving for an estimate of $\Delta\boldsymbol{x}$ as $\boldsymbol{x}_{true} = \boldsymbol{x}_0 + \Delta\boldsymbol{x}$,

$$\widehat{\Delta\boldsymbol{x}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\Delta\boldsymbol{\rho}. \tag{21}$$

In order to obtain a more accurate estimate, $\widehat{\Delta\boldsymbol{x}}$ is added to $\boldsymbol{x}_0$ to become the new value of $\boldsymbol{x}_0$:

$$\boldsymbol{x}_{0_{new}} = \boldsymbol{x}_{0_{old}} + \widehat{\Delta\boldsymbol{x}} \tag{22}$$

Equations (20 - 22) are then recalculated with the updated $\boldsymbol{x}_0$. Through each iteration, the values of $\boldsymbol{x}_{0_{new}}$ theoretically converge to estimate the value of $\boldsymbol{x}$. However, the initial $\boldsymbol{x}_0$ may not be close enough to the real value of $\boldsymbol{x}$ to allow for a valid linearization, or the cost surface created may have local maxima to which the LSE may converge. An initial $\boldsymbol{x}_0$ that is too inaccurate may cause a divergent solution to $\boldsymbol{x}$. If the LSE solution produces a convergent result, the last calculated $\boldsymbol{x}_0$ is assigned as $\hat{\boldsymbol{x}}$, the estimate of the state vector. In Tier 1, all tests produced solutions that converged to the correct locations.

3.2.8 Methodology Summary.

The methodology for Tier 1 can be summarized through Figure 3. First, the true mobile microphone and speaker locations are determined using the Vicon system. The audio from each speaker is then sequentially played while the mobile microphone records for one hundred trials. The single audio file is then separated into trials. Each recorded impulse in the trial is then attributed to its corresponding speaker based on the order in which it was recorded. Using the known speaker locations and timing of the impulses, TOA estimates are formed. Due to receiver clock error, the TOA

measurements must be corrected for drift. The measurements are then normalized to offset quantization error. Finally, an iterative LSE process produces the estimated mobile microphone location. The same general method of microphone positioning is used in later tiers, with changes compensating for fewer assumptions on the signals and environment.

3.3 Tier 1 Results

3.3.1 Methods for Data Characterization.

Measurement Domain Analysis.

For each microphone test location, the TOA range measurements were compared to the true distances from the microphone to each speaker, and the results are presented in histograms (Figure 11 is one example). Positive values indicate the distances derived from the truth data were greater than the distances estimated via sound whereas negative values indicate the distances derived from the truth data were less than the distances estimated via sound.

Error Ellipses.

A confidence error ellipse characterizes the distribution of location estimates. The analyses of results for all tiers use 95% confidence error ellipses. If a test were to be repeated again with 100 trials, 95 of the 100 location estimates should be expected to fall within the 95% confidence error ellipse of that testpoint. The center of the ellipse is the mean location of the estimates. The covariance of the estimates in the x and y dimensions determines the size and eccentricity of the ellipse. As the overall variance of the data increases, the size of the ellipse increases. There is often more variance in one dimension than the other, in which case, the ellipse is more eccentric and orients lengthwise in the direction of greater variance.

Position Error.

Mean values and standard deviations for error in the x and y directions are reported for each testpoint as well as the Distance Root Mean Squared (DRMS). DRMS is a single value metric that quantifies the estimation accuracy and precision of a particular testpoint and allows for comparison to other testpoints. If the n^{th} solution is represented as $[\hat{x}_{m_n}, \hat{y}_{m_n}]^T$ and the true position is $[x_m, y_m]^T$, then the DRMS of N solutions is calculated as

$$\text{DRMS} = \sqrt{\frac{\sum_{n=1}^{N} \left((\hat{x}_{m_n} - x_m)^2 + (\hat{y}_{m_n} - y_m)^2 \right)}{N}}, \tag{23}$$

DRMS accounts for both the bias of the estimates and variance between the estimates of each trial.

Clock Error.

The LSE estimation not only solves for the location of the mobile microphone, but also the estimated clock error of the receiver for the TOA or TDOA measurements. For TOA measurements in Tier 1, the clock error value itself does not provide useful information, since the offset is relative to an arbitrary point in time, as mentioned in Section 3.2.4. Even with an arbitrary time offset, the LSE solution provided consistent clock error estimates with an approximately normal distribution for all tests. So, the standard deviation of the clock error is reported to demonstrate the consistency of the timing estimate for use in applications with a non-arbitrary reference time. For Tiers 2 and beyond, which use TDOA measurements, both clock error estimate means and standard deviations are reported to demonstrate the accuracy of the estimated state vector. Because clock errors cancel out in TDOA measurements from the same receiver, the true time offset is considered to be zero.

Comparison to Dilution of Precision.

As discussed in Section 2.3.1, DOP calculations quantify how well location estimates can be made from the measured TOA values. DOP is only dependent on the configuration of the microphone and sound sources relative to one another. With fixed sound source locations, Figure 9 visualizes how well the solution may be expected to work for any mobile microphone location in the testing area. Because DOP does not change in Tier 2 by introducing a single reference microphone, Figure 9 may also be referenced for Tier 2 tests. Table 3 contains the DOP values at each of the testpoints for Tiers 1 and 2.

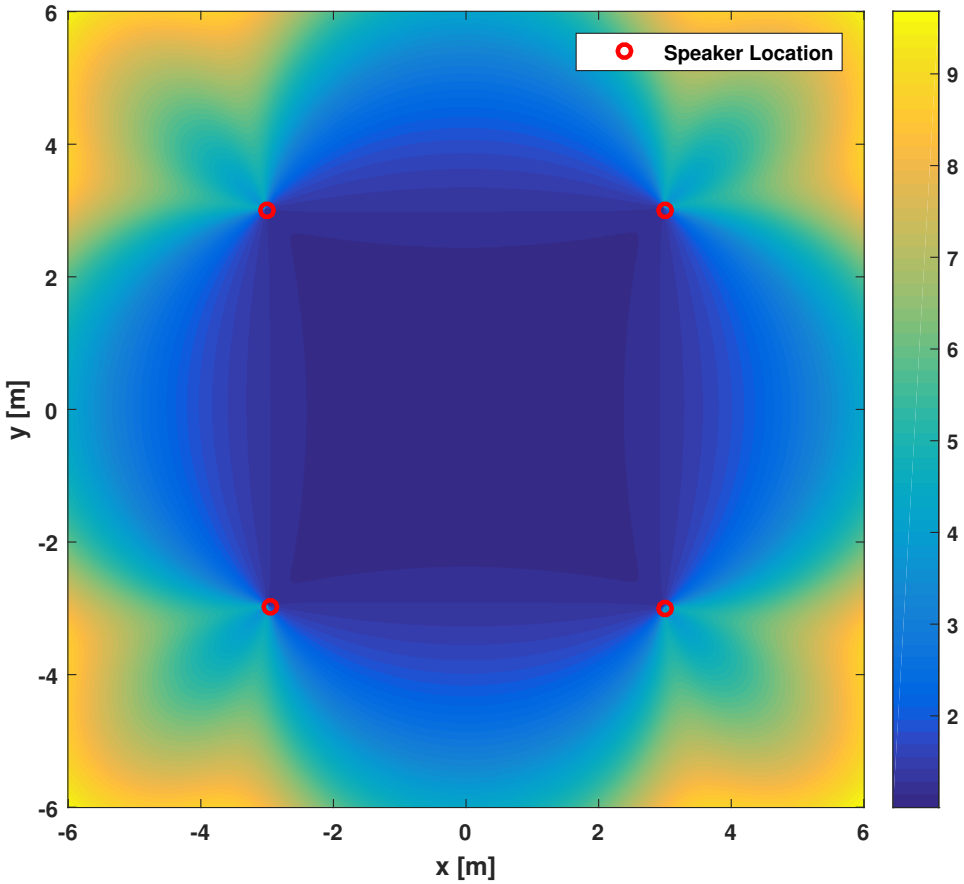


Figure 9. DOP map of testbed for Tier 1 and 2 tests. Color corresponds to DOP as a function of the location of the mobile microphone.

3.3.2 Results by Testpoint.

Figure 10 shows the location of each testpoint in the testing area. Because all tiers use the same testpoints, Figure 10 may be used as reference for all further tiers. Table 3 provides summary results for Tier 1 at each testpoint.

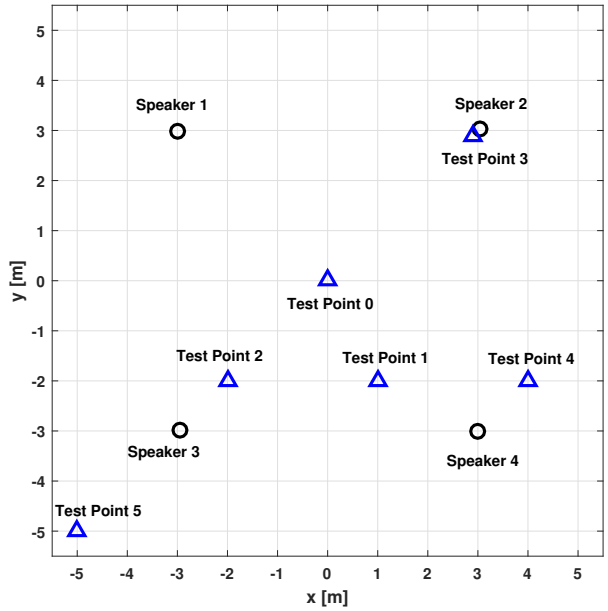


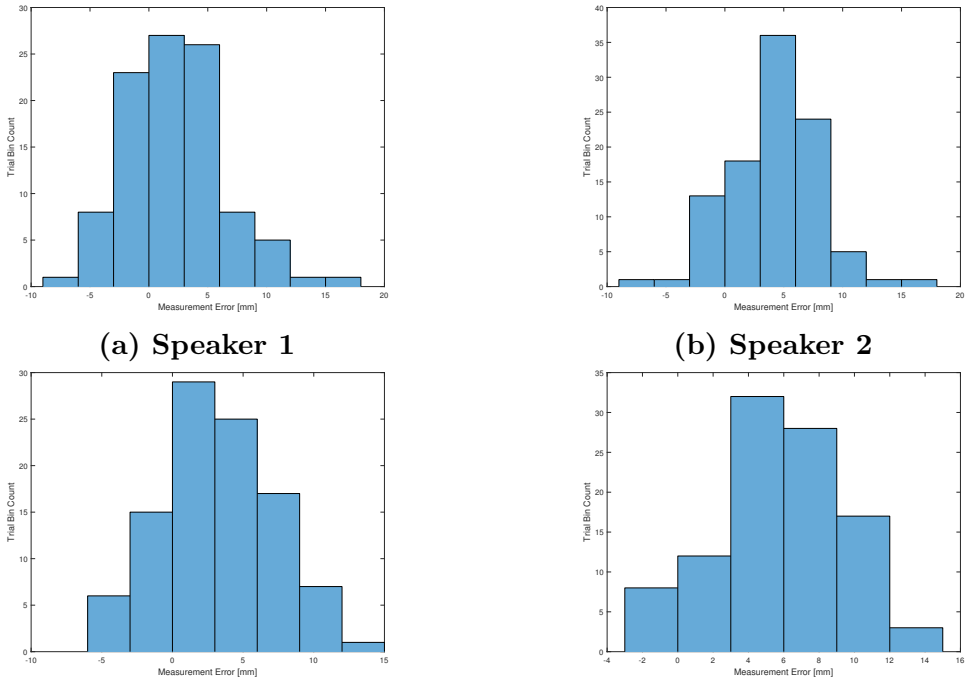
Figure 10. Locations of sound sources and testpoints for the mobile microphone

Table 3. Results for Test Performed in Tier 1

Testpoint	Approx. Coords.	DOP	Clock Error (μs)	Position Error (cm)				
			$\sigma_{\delta t}$	$\bar{x}$	σ_x	$\bar{y}$	σ_y	DRMS
0	(0,0)	1.000	39.98	0.17	0.28	-0.09	0.19	0.39
1	(1,-2)	1.069	35.31	-0.98	0.30	-0.23	0.17	1.06
2	(-2,-2)	1.075	23.74	0.91	0.21	-0.06	0.18	0.95
3	(2.9,2.9)	1.171	17.38	0.95	0.21	-0.28	0.33	1.07
4	(4,-2)	2.387	56.04	0.25	0.47	0.90	0.21	1.06
5	(-5,-5)	7.149	144.13	-1.99	-2.81	1.29	1.48	3.96

Testpoint 0.

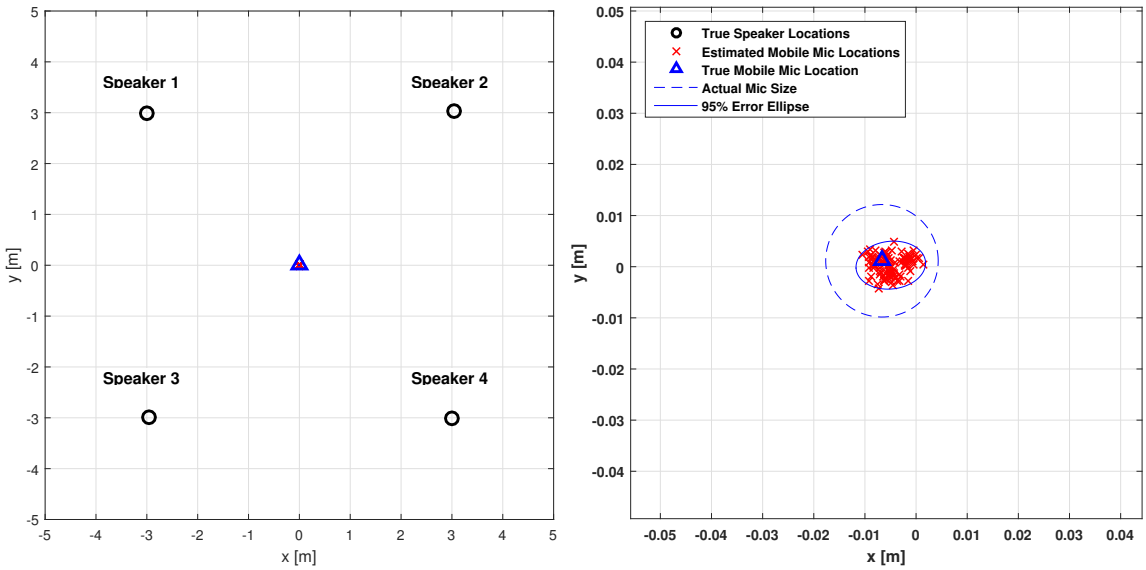
Testpoint 0 is located at the center of the test area, equidistant from each sound source as shown in Figure 12a. With the central location of the microphone, the speed of sound computed in Equation (8) does not adversely affect the location estimate. As shown in Figure 11, there is a slight positive bias on the order of 3-5 mm in the distribution of error for each distance estimate between the mobile microphone and each speaker. A similarly ordered bias is seen in Figure 12b with the location estimation with estimates averaging slightly right of the true location. All estimates fall within the area of the microphone diaphragm.



(a) Speaker 1 **(b) Speaker 2**

(c) Speaker 3 **(d) Speaker 4**

Figure 11. TOA measurement error distribution at Testpoint 0

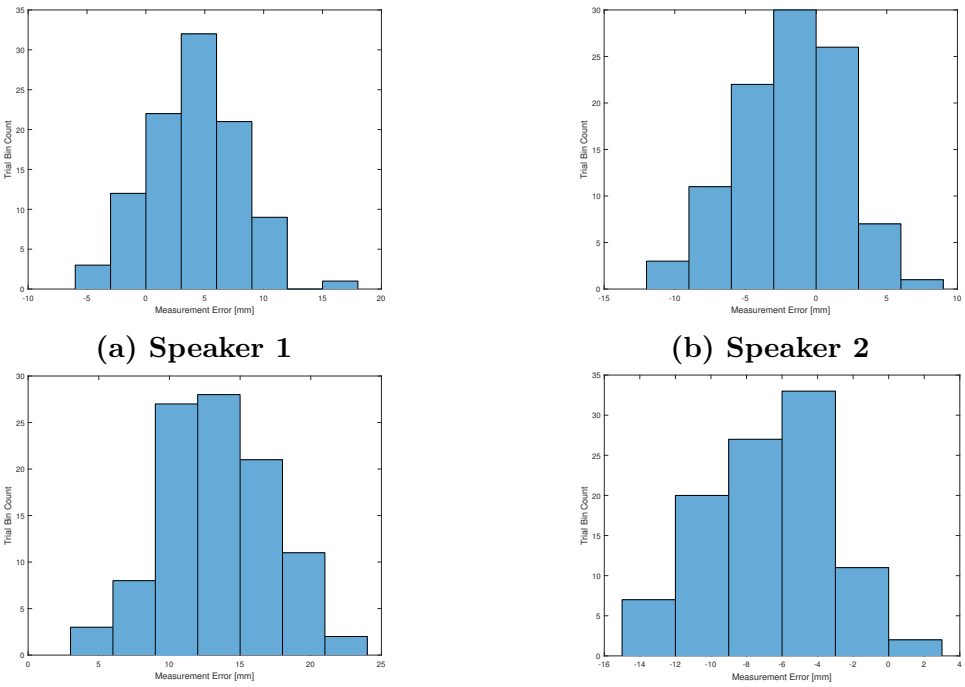


(a) Full test area **(b) Magnified at Testpoint 0**

Figure 12. Estimated location of mobile microphone at Testpoint 0

Testpoint 1.

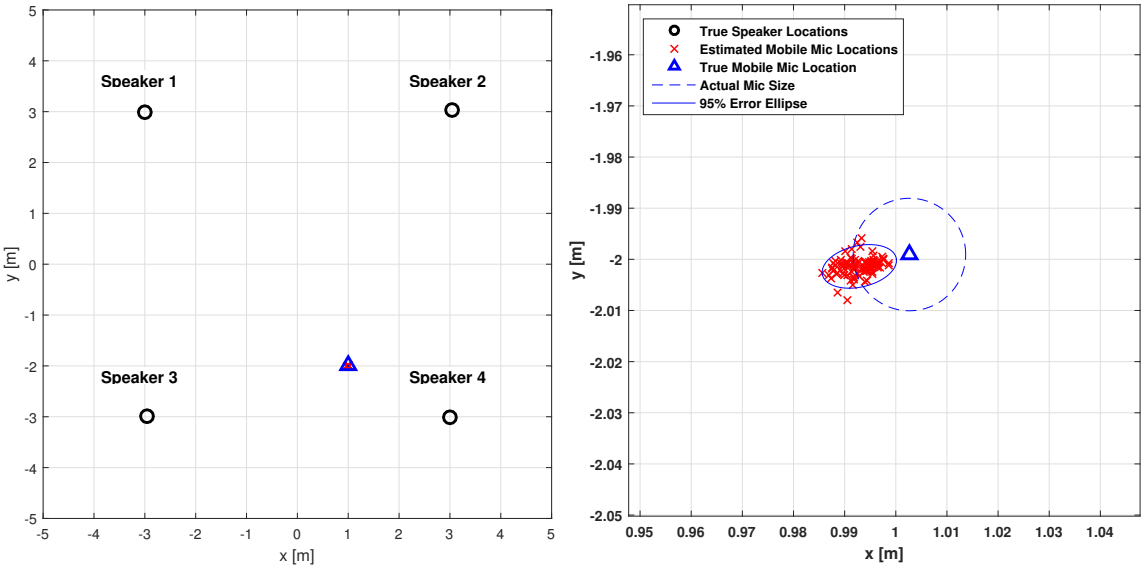
As seen in Figure 14a, none of the distances from the microphone to speaker are the same at Testpoint 1, where errors in speed of sound model or similar errors could affect the accuracy of TDOA measurements. Figure 13 shows inconsistency in the distance measurement errors between sound sources, with Speakers 1 and 3 showing positive bias, speaker 2 nearly unbiased, and Speaker 4 showing negative bias. The inconsistency of the biases affects the LSE estimation by moving the location estimates to the left and slightly below the true mobile microphone position. Figure 14b shows approximately half of the estimates within the area of the microphone diaphragm.



(a) Speaker 1 (b) Speaker 2

(c) Speaker 3 (d) Speaker 4

Figure 13. TOA measurement error distribution at Testpoint 1

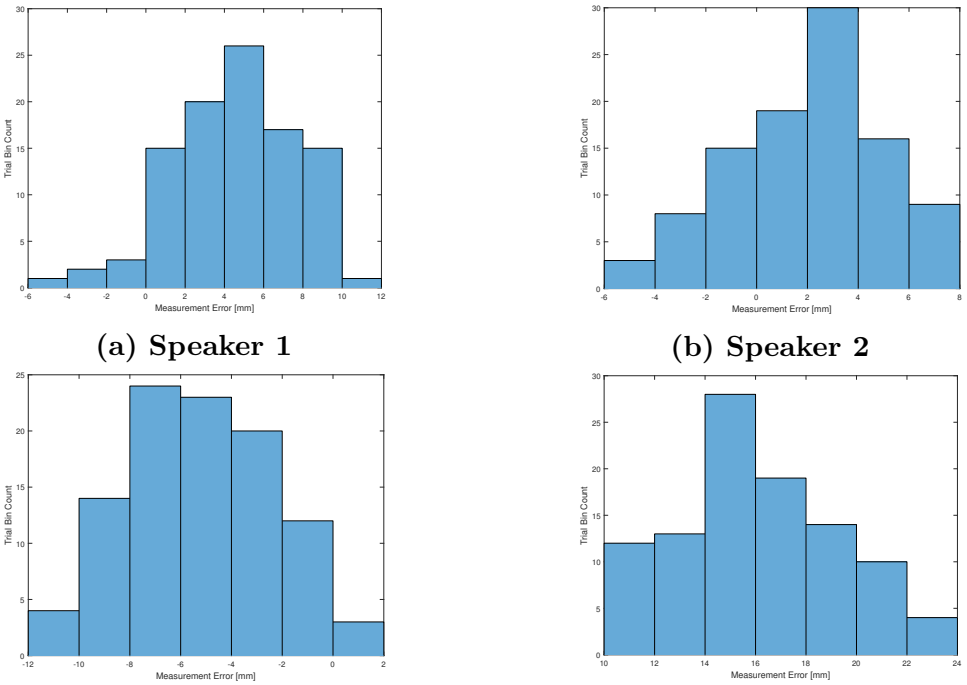


(a) Full test area (b) Magnified at Testpoint 1

Figure 14. Estimated location of mobile microphone at Testpoint 1

Testpoint 2.

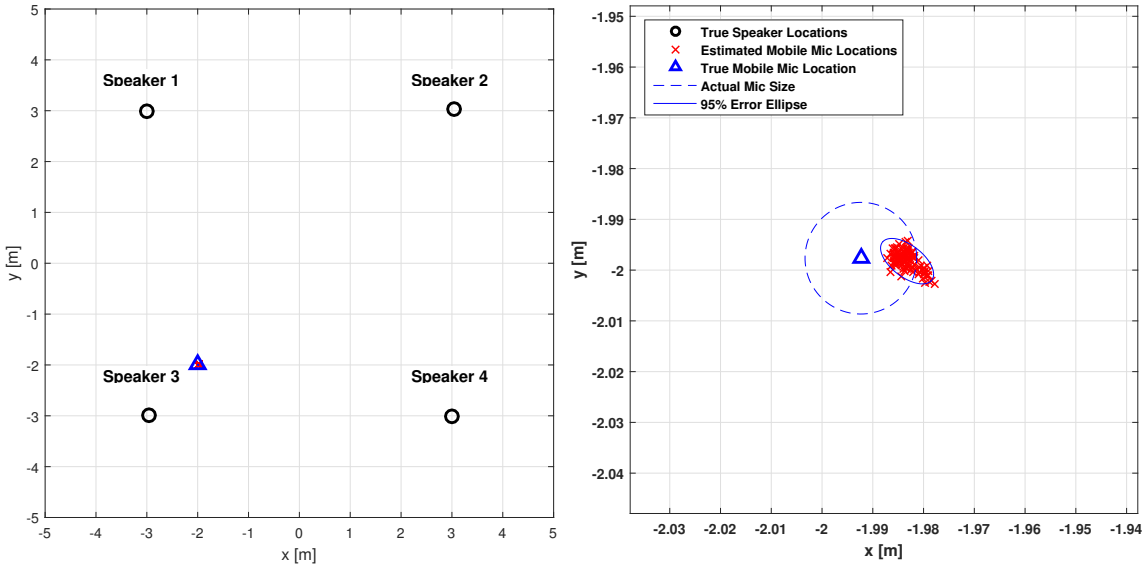
As shown in Figure 16a, Testpoint 2 is equidistant from Speakers 1 and 4, and co-linear with Speakers 2 and 3. Because of the equidistance between Speakers 1 and 4, similar biases in the measurement domain were expected. However, as shown in Figure 15, the bias for Speaker 1 was around 4 mm, but Speaker 4 showed a bias around 16 mm. The location estimates pulled toward Speaker 4, possibly due to the high measurement bias. Over half of the estimates were contained within the area of the mobile microphone diaphragm as shown in Figure 16b.



(a) Speaker 1 **(b) Speaker 2**

(c) Speaker 3 **(d) Speaker 4**

Figure 15. TOA measurement error distribution at Testpoint 2

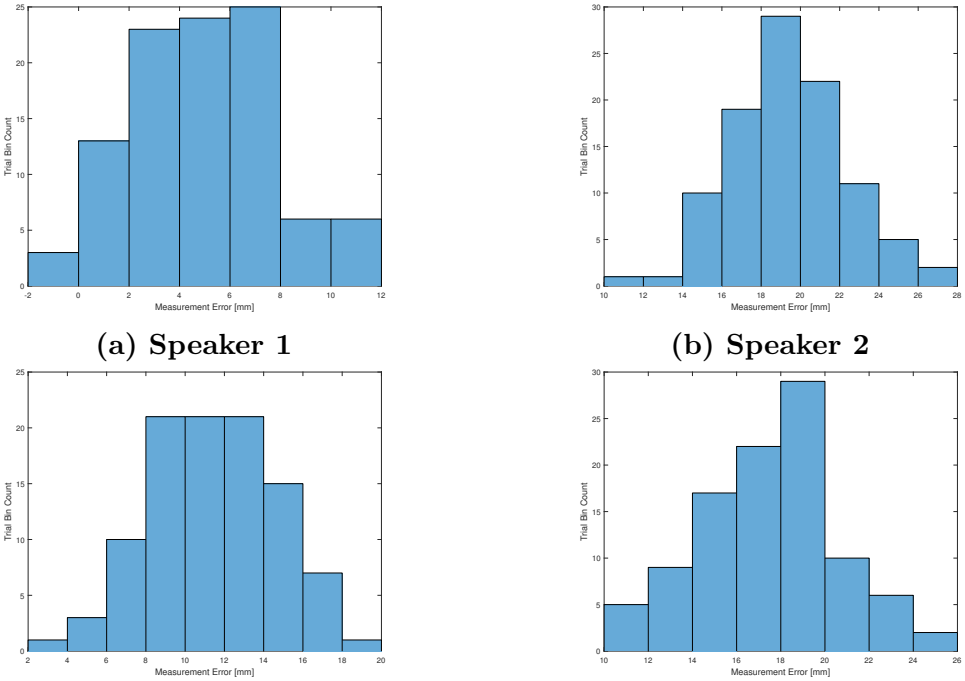


(a) Full test area **(b) Magnified at Testpoint 2**

Figure 16. Estimated location of mobile microphone at Testpoint 2

Testpoint 3.

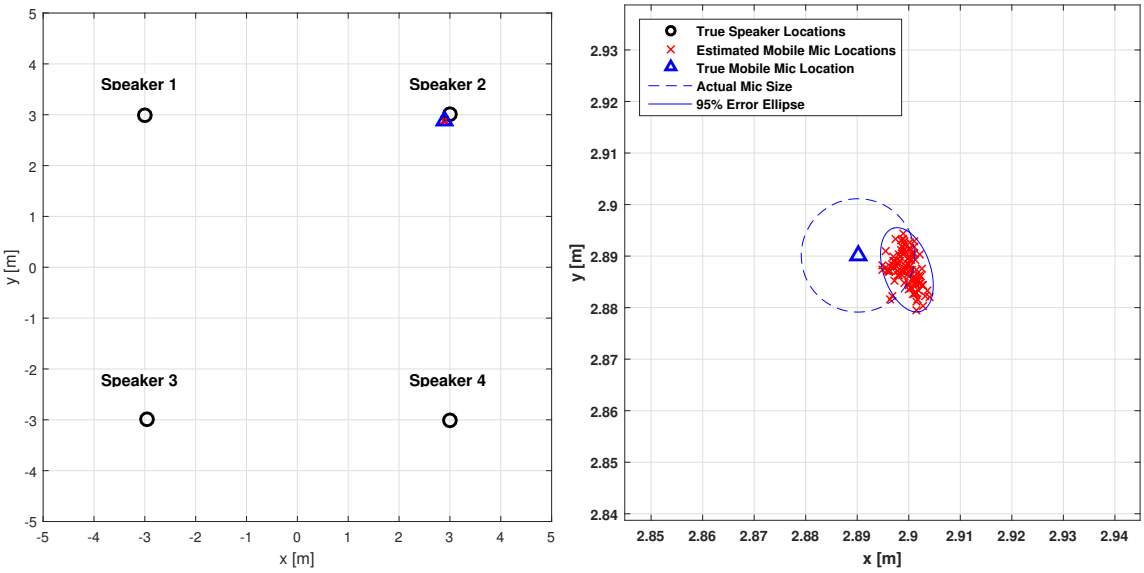
In order to test the effects of close proximity of the mobile microphone to a sound source, Testpoint 3 is located adjacent to Speaker 2, as shown in Figure 18a. Slight measurement biases were present from Speakers 2, 3, and 4 as shown in Figure 17. These biases caused the location estimates to move right and slightly below the true microphone location. Approximately half of the estimates were contained within the area of the mobile microphone diaphragm as shown in Figure 18b.



(a) Speaker 1 (b) Speaker 2

(c) Speaker 3 (d) Speaker 4

Figure 17. TOA measurement error distribution at Testpoint 3

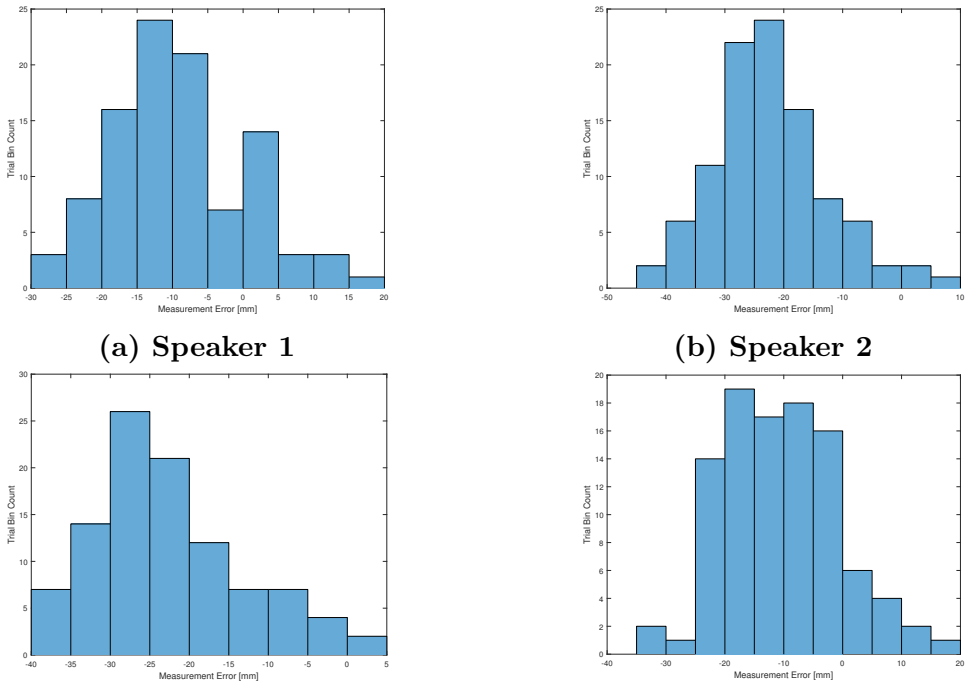


(a) Full test area (b) Magnified at Testpoint 3

Figure 18. Estimated location of mobile microphone at Testpoint 3

Testpoint 4.

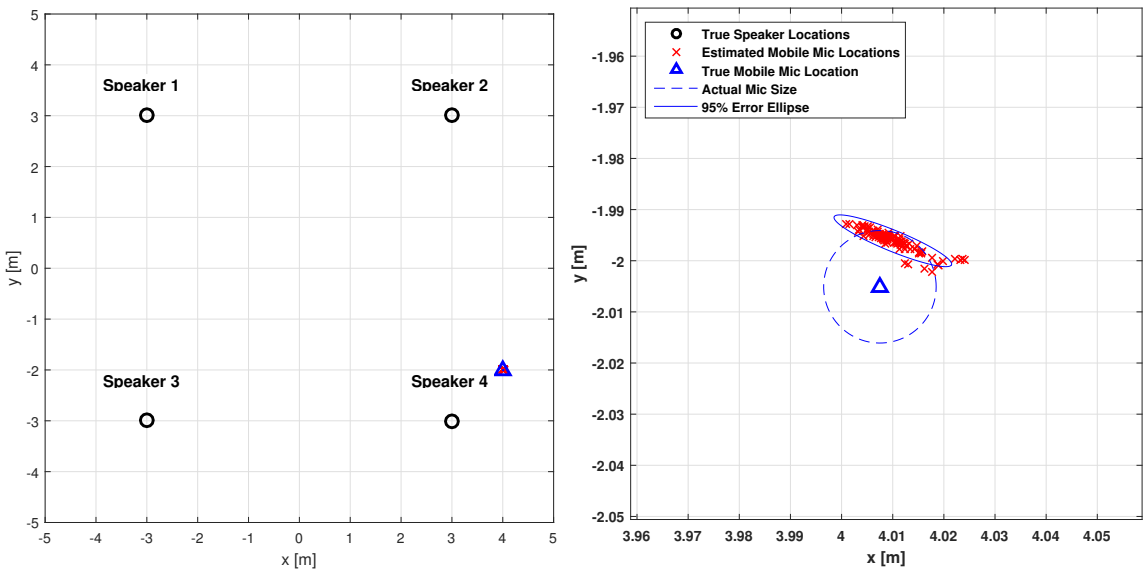
Testpoint 4 is outside of the perimeter of the four speakers, as shown in Figure 20a, where the DOP is significantly greater compared to values inside the speaker perimeter. The effects of increased DOP are shown through the increased size of the 95% confidence error ellipse in Figure 20b. Also of note is the orientation of the ellipse towards the center of the test area.



(a) Speaker 1 **(b) Speaker 2**

(c) Speaker 3 **(d) Speaker 4**

Figure 19. TOA measurement error distribution at Testpoint 4

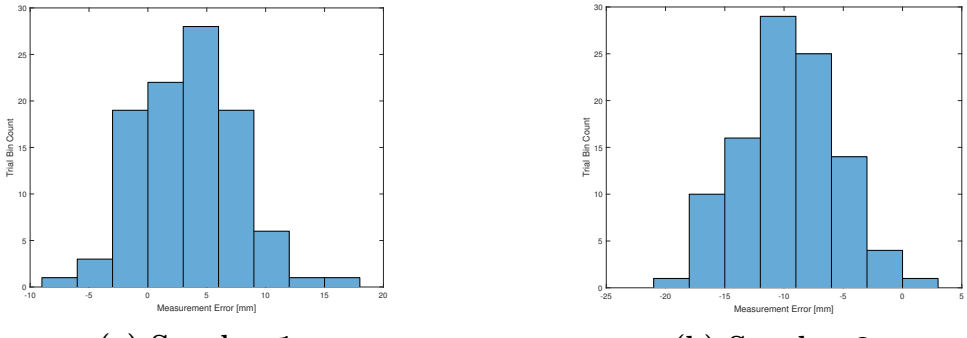


(a) Full test area **(b) Magnified at Testpoint 4**

Figure 20. Estimated location of mobile microphone at Testpoint 4

Testpoint 5.

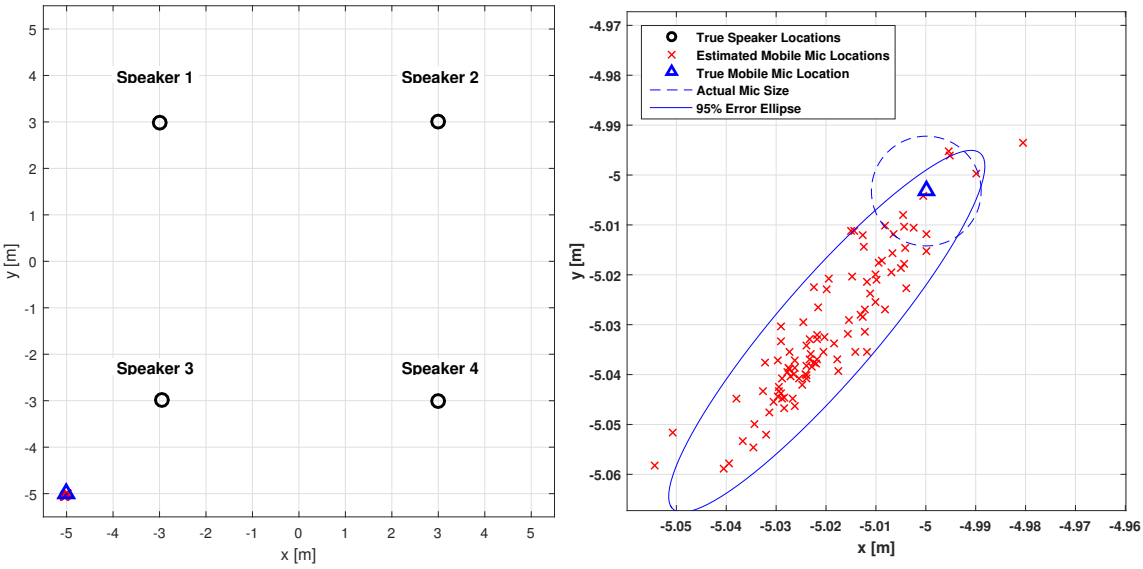
Figure 22a shows testpoint 5 is far outside of the speaker perimeter, in an area with much higher DOP relative to the other test locations. While the biases in the measurement domain shown in Figure 21 were similar to biases present in Testpoints 1-4, the larger DOP magnified the bias in the location estimation to be much greater than previous tests with lower DOP. As shown in Figure 22b, estimates were biased to the left and below the true microphone location. The large variance in estimates led to an error ellipse much greater in area than seen in previous testpoints. The DRMS for Testpoint 5 was 4.04 cm, also significantly larger than previous testpoints.



(a) Speaker 1 (b) Speaker 2

(c) Speaker 3 (d) Speaker 4

Figure 21. TOA measurement error distribution at Testpoint 5



(a) Full test area (b) Magnified at Testpoint 5

Figure 22. Estimated location of mobile microphone at Testpoint 5

3.4 Tier 2 Methodology

3.4.1 Section Overview.

In Tier 2, the signals are still emitted sequentially, but the times that the signals are emitted are no longer known nor consistent. Because exact signal timing is not known, TDOA measurements must be used instead of just TOA measurements. The methodology for Tier 2 is summarized in Figure 23. New methods introduced include TDOA estimation, location estimation through closed form solutions, and TDOA LSE location estimation. Drift correction was a necessary correction to TOA estimates in Tier 1 because of receiver clock error. However, with the introduction of TDOA estimates, receiver clock error cancels out, making drift correction unnecessary. With the exception of methods discussed in this section, all testing was performed in the same manner as Tier 1.

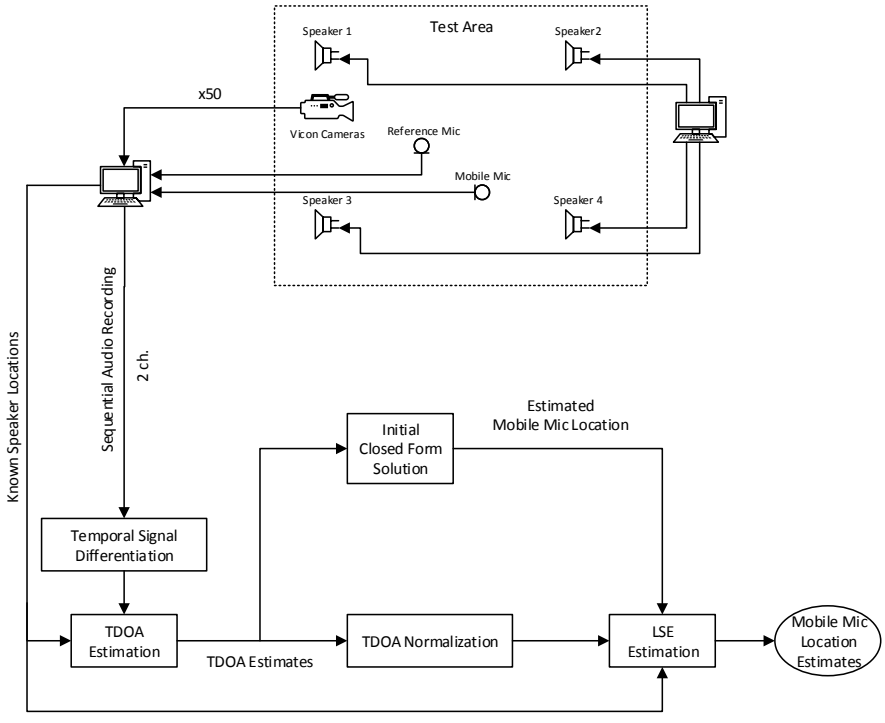


Figure 23. Methodology for obtaining mobile microphone location estimates in Tier 2.

3.4.2 Time Difference of Arrival Estimation.

The peaks in both the mobile and reference microphone signals are measured in the same way as Tier 1, where $\mathbf{T}_m$ represents the time of the peaks detected from the mobile microphone and $\mathbf{T}_r$ represents the time of the peaks detected from the reference microphone. TDOA measurements are formed by

$$\Delta\mathbf{T}_{mr} = \mathbf{T}_m - \mathbf{T}_r = \begin{bmatrix} T_m^{s_1} \\ \vdots \\ T_m^{s_K} \end{bmatrix} - \begin{bmatrix} T_r^{s_1} \\ \vdots \\ T_r^{s_K} \end{bmatrix}. \tag{24}$$

where superscripts denote the respective sound source. The clocks measuring the TOAs from the mobile microphone and the reference microphone may be separate, and in practical scenarios, will be. However, the clock error between the mobile receiver clock and reference receiver clock is estimated as part of the LSE solution, and does not affect position estimate accuracy. Multiplying TDOA measurements by the speed of sound forms the range difference measurements of the trial:

$$\Delta\boldsymbol{\rho}_{mr} = c\Delta\mathbf{T}_{mr} = \begin{bmatrix} \rho_{mr}^{s_1} \\ \vdots \\ \rho_{mr}^{s_K} \end{bmatrix} \tag{25}$$

3.4.3 Location Estimation Through Closed Form Solutions.

With the introduction of TDOA measurements, a closed form solution to $\hat{\boldsymbol{x}}$ is possible [19]. In the absence of measurement error, the closed form solution produces the exact location. However, measurement errors generally cause the closed form solution to be less accurate than a convergent LSE estimation. The closed form solution is still useful, because it is often accurate enough to provide an initial $\boldsymbol{x}_0$ that leads to a convergent result from the LSE estimation, instead of initializing $\boldsymbol{x}_0$

to the center of the test area, as done in Tier 1.

The closed form solution to the mobile microphone location is given by [19] as

$$\hat{\boldsymbol{x}}_{cf} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T(\mathbf{z} - \boldsymbol{\rho}\hat{R}_s) + \begin{bmatrix} x_{s_1} \\ y_{s_1} \end{bmatrix}, \tag{26}$$

where $\hat{\boldsymbol{x}}_{cf}$ is the closed form solution of the mobile microphone position and S is the regressor matrix of sound source locations relative to the first sound source such that

$$\mathbf{S} \triangleq \begin{bmatrix} x_{s_2} - x_{s_1} & y_{s_2} - y_{s_1} \\ \vdots & \vdots \\ x_{s_K} - x_{s_1} & y_{s_K} - y_{s_1} \end{bmatrix}, \tag{27}$$

$\mathbf{z}$ is a vector such that

$$\mathbf{z} \triangleq \frac{1}{2}\begin{bmatrix} (x_{s_2} - x_{s_1})^2 + (y_{s_2} - y_{s_1})^2 - (\Delta\rho_{mr}^{s_2} - \Delta\rho_{mr}^{s_1})^2 \\ \vdots \\ (x_{s_K} - x_{s_1})^2 + (y_{s_K} - y_{s_1})^2 - (\Delta\rho_{mr}^{s_K} - \Delta\rho_{mr}^{s_1})^2 \end{bmatrix}. \tag{28}$$

$$\hat{R}_s = \boldsymbol{\rho}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-2}\mathbf{S}^T\mathbf{z} \pm \frac{\sqrt{[\boldsymbol{\rho}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-2}\mathbf{S}^T\mathbf{z}]^2 + \mathbf{z}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-2}\mathbf{S}^T\mathbf{z} \cdot [1 - \boldsymbol{\rho}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-2}\mathbf{S}^T\boldsymbol{\rho}]}}{\boldsymbol{\rho}^T\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-2}\mathbf{S}^T\boldsymbol{\rho} - 1}. \tag{29}$$

In the above equations, the subscripts $s_1, \ldots, s_K$ represent the speaker number. When at least three of the sound sources are not co-linear, then the matrix $\mathbf{S}$ has full rank. However, it is possible for Equation (29) to have imaginary roots, such that the solution to $\hat{R}_s$ and the location of the mobile microphone cannot be determined [19]. In such a case, $\boldsymbol{x}_0$ would have been initialized to the center of the test area. If the closed form solution produced a valid, real-valued estimate, $\boldsymbol{x}_0$ was initialized as $\hat{\boldsymbol{x}}_{cf}$ for the LSE. In all tests, the closed form solution produced a valid estimate.

3.4.4 Least Squares Location Estimation.

The range difference measurement for the k^{th} sound source relative to the mobile and reference microphones is

$$\Delta\rho_{mr}^{s_k} = \rho_m^{s_k} - \rho_r^{s_k} + \upsilon_k \tag{30}$$

where $\rho_r^{s_k}$ is the range from the k^{th} sound source to the reference microphone and $\rho_m^{s_k}$ is the range from the k^{th} sound source to the mobile microphone such that

$$\Delta\rho_{mr}^{s_k} = \left(\sqrt{(x_{s_k} - x_m)^2 + (y_{s_k} - y_m)^2} + \delta t_m\right) - \left(\sqrt{(x_{s_k} - x_r)^2 + (y_{s_k} - y_r)^2} + \delta t_r\right) + \upsilon_k. \tag{31}$$

The difference between the mobile clock error, δt_m, and the reference clock error, δt_r is given as δt, simplifying Equation (31) to

$$\Delta\rho_{mr}^{s_k} = \sqrt{(x_{s_k} - x_m)^2 + (y_{s_k} - y_m)^2} - \sqrt{(x_{s_k} - x_r)^2 + (y_{s_k} - y_r)^2} + \delta t + \upsilon_k. \tag{32}$$

In testing, the mobile and reference audio were recorded using the same clock, so δt is expected to be 0. However, in scenarios were the mobile and reference audio are recorded on separate systems, δt is the difference in time between the mobile and reference clocks.

Equation (32) may be further simplified as

$$\Delta\rho_{mr}^{s_k} = h(\boldsymbol{x}) + \upsilon_k, \tag{33}$$

$$h(\boldsymbol{x}) = \sqrt{(x_{s_k} - x_m)^2 + (y_{s_k} - y_m)^2} - \sqrt{(x_{s_k} - x_r)^2 + (y_{s_k} - y_r)^2} + \delta t. \tag{34}$$

Represented in matrix form for range measurements from all sound sources,

$$\boldsymbol{\Delta\rho}^{s}_{mr} = \mathbf{h}(\boldsymbol{x}) + \boldsymbol{v}. \tag{35}$$

A process similar to the LSE algorithm outlined in Tier 1 was then implemented in order to produce $\hat{\boldsymbol{x}}$ from TDOA measurements.

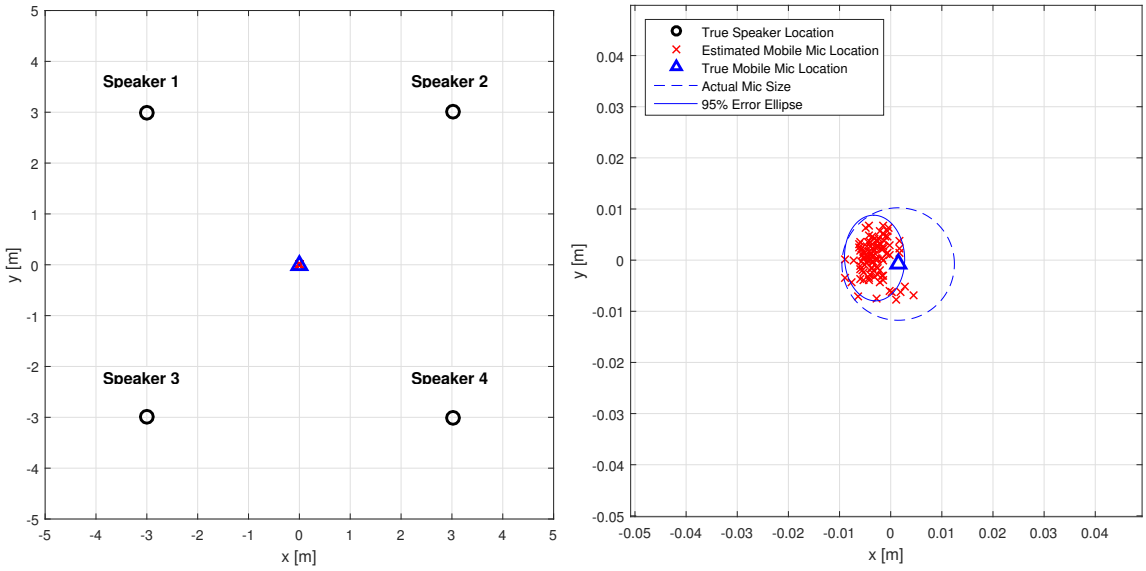
3.5 Tier 2 Results

Results are presented for select testpoints in Tier 2 in this section. Because results produced similar findings through measurement domain analysis to Tier 1, measurement domain results are not discussed in depth for any of the remaining tiers. As mentioned in Section 3.3.1, both the mean value and variance of the time offset are included in the results of Tier 2 and all following tiers.

Table 4. Results for Test Performed in Tier 2

Testpoint	Approx. Coords.	DOP	Clock Error (μs)		Position Error (cm)				
			δt	$\sigma_{\delta t}$	$\bar{x}$	σ_x	$\bar{y}$	σ_y	DRMS
0	(0,0)	1.000	23.39	10.73	-0.46	0.24	0.12	0.34	0.63
1	(1,-2)	1.069	31.22	5.15	-0.56	0.34	-0.53	0.43	0.94
2	(-2,-2)	1.075	48.29	8.18	1.16	0.38	0.46	0.43	1.37
3	(2.9,2.9)	1.171	25.21	6.24	-0.45	0.17	0.488	0.21	0.71
4	(4,-2)	2.387	32.02	22.60	1.11	0.72	-0.70	0.42	1.56
5	(-5,-5)	7.149	82.97	49.96	0.54	1.23	1.24	1.48	2.34

Testpoint 0.

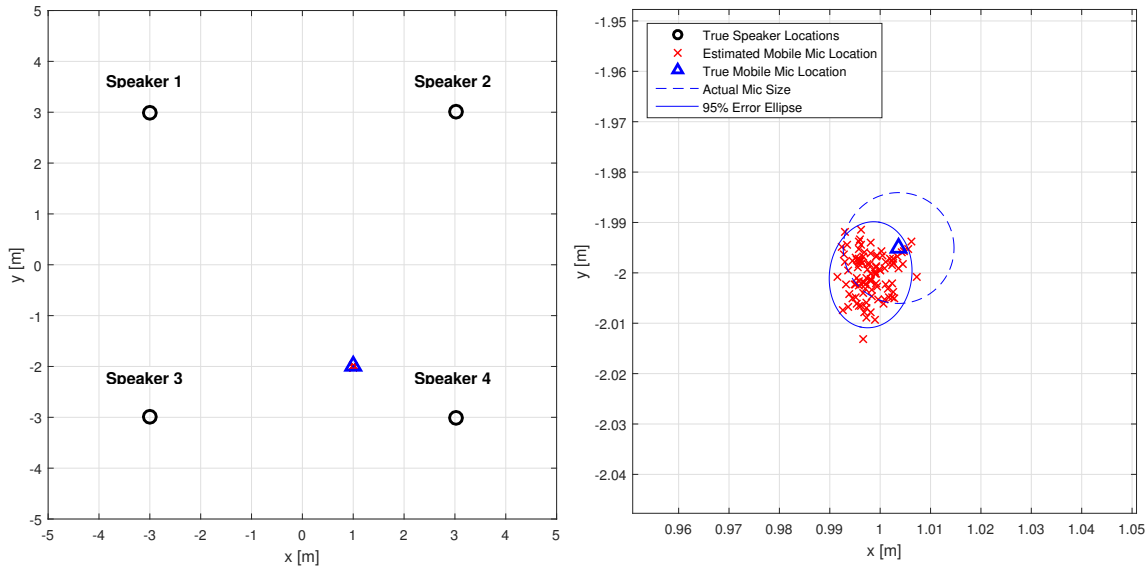


(a) Full test area **(b) Magnified at Testpoint 0**

Figure 24. Estimated location of mobile microphone at Testpoint 0

The test configuration and results for Testpoint 0 are shown in Figure 24. The results for Testpoint 0 show TDOA measurements when signal timing is not known (Tier 2) provide slightly different results to using TOA measurements when timing is known as in Tier 1. While the bias is of similar magnitude, estimates are slightly left of the true microphone location. All estimates fell within the area of the microphone diaphragm as shown in Figure 24b. The DRMS increased almost twofold to 0.63 cm. Because TDOA is a difference of two TOA measurements, overall measurement noise is expected to increase, leading to higher positioning error despite equivalent DOPs seen in Tier 1.

Testpoint 1.

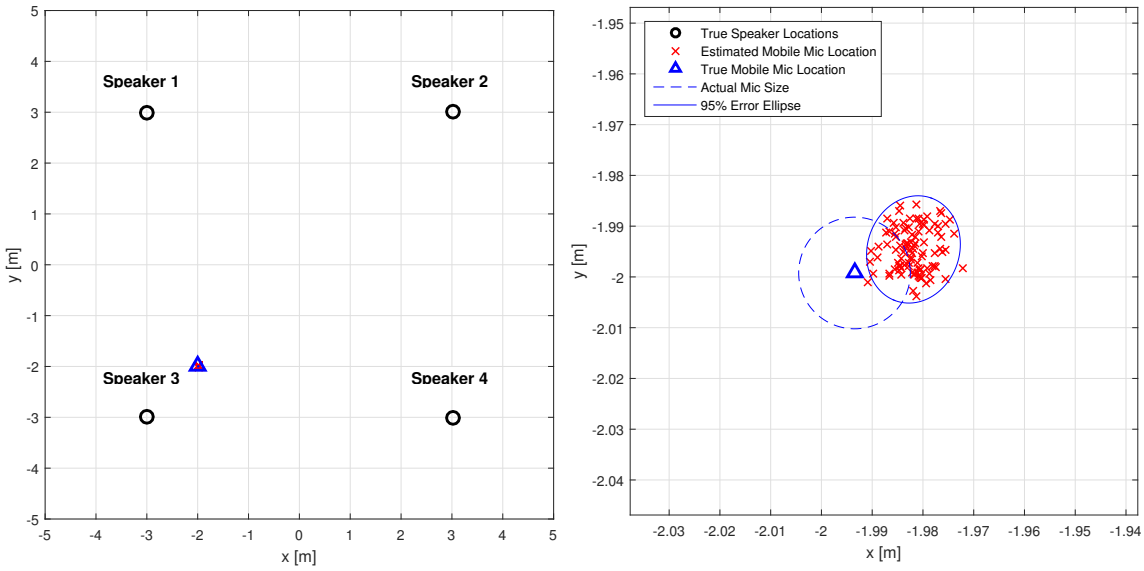


(a) Full test area **(b) Magnified at Testpoint 1**

Figure 25. Estimated location of mobile microphone at Testpoint 1

The test area configuration and results for Testpoint 1 are shown in Figure 25. The bias was similar between Tiers 1 and 2 for Testpoint 1, pulling slightly below and left of the true microphone location. However, the shape of the error ellipse in Tier 2 is slightly more circular than the error ellipse shown for Tier 1. The majority of estimates still fell within the area of the microphone diaphragm as shown in Figure 25b. The DRMS was 0.94 cm, slightly less than the DRMS of Tier 1 for Testpoint 1.

Testpoint 2.

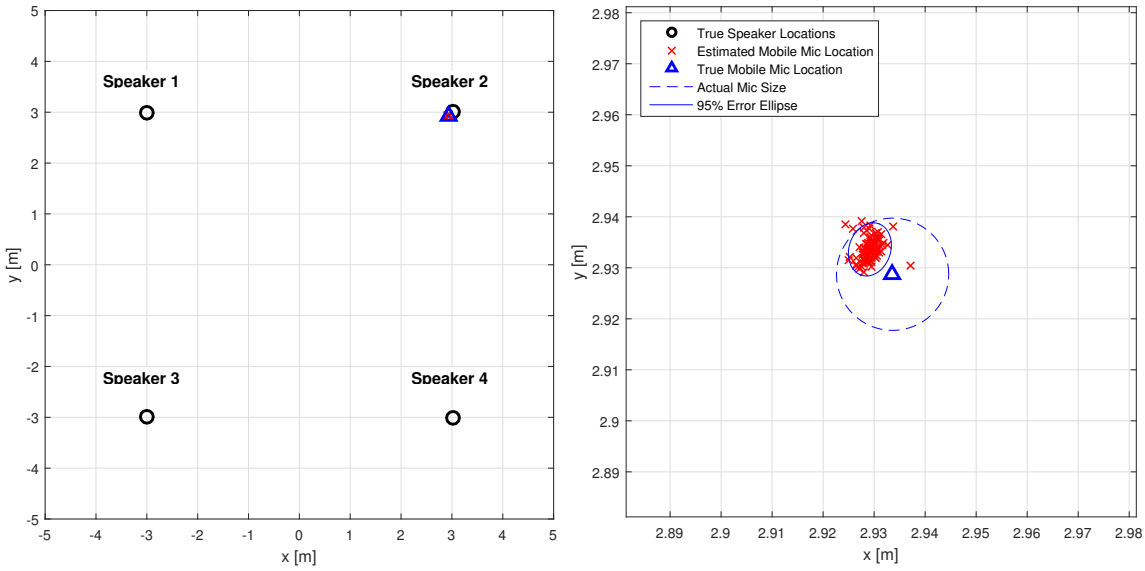


(a) Full test area **(b) Magnified at Testpoint 2**

Figure 26. Estimated location of mobile microphone at Testpoint 2

Figure 26 shows the test area layout and results for Testpoint 2. Approximately one third of the estimates fell within the area of the microphone diaphragm as shown in Figure 26b. While the bias of the estimates was similar to results from Tier 1, the spread of the estimates was slightly larger, causing the DRMS to increase by 42 mm to 1.37 cm. The the mean clock error was 48 μs with a standard deviation of 8.18 μs, one third the size of the standard deviation from Tier 1.

Testpoint 3.

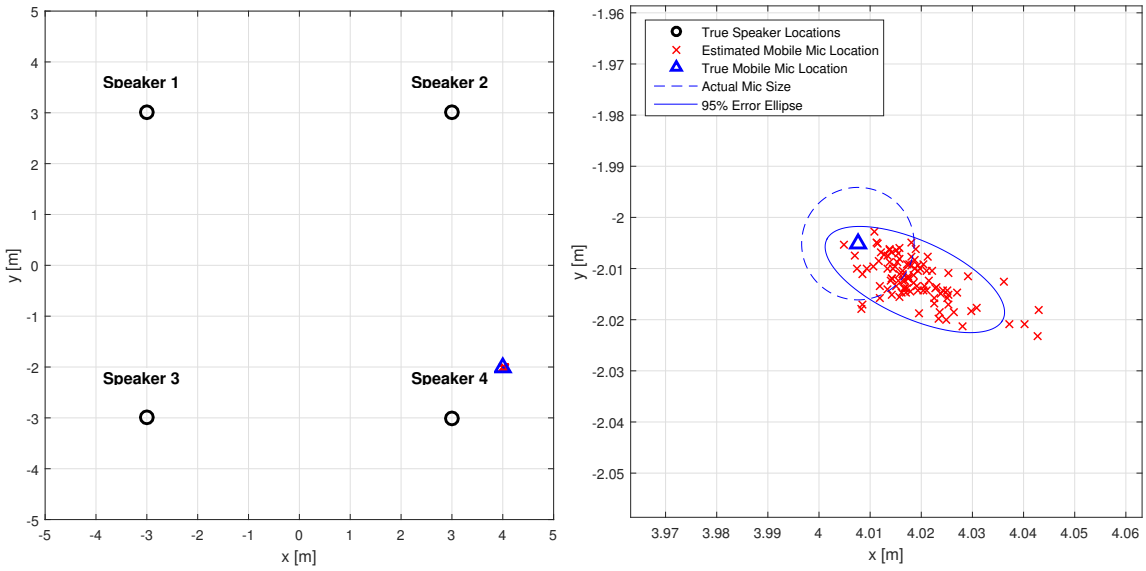


(a) Full test area **(b) Magnified at Testpoint 3**

Figure 27. Estimated location of mobile microphone at Testpoint 3

The accuracy of TDOA measurements was not altered by the close proximity of Speaker 2 to the mobile microphone, as seen in Figure 27a. Both the bias and variance of the estimates were much less than comparable results from Tier 1, leading to a DRMS of 0.71 cm. As seen in Figure 27b, 97 of the 100 estimates fell within the area of the microphone diaphragm. The accurate results for Tier 2 at this testpoint form a basis of comparison for Tiers 4 and 5, where the accuracy is degraded due to changes in TDOA acquisition.

Testpoint 4.

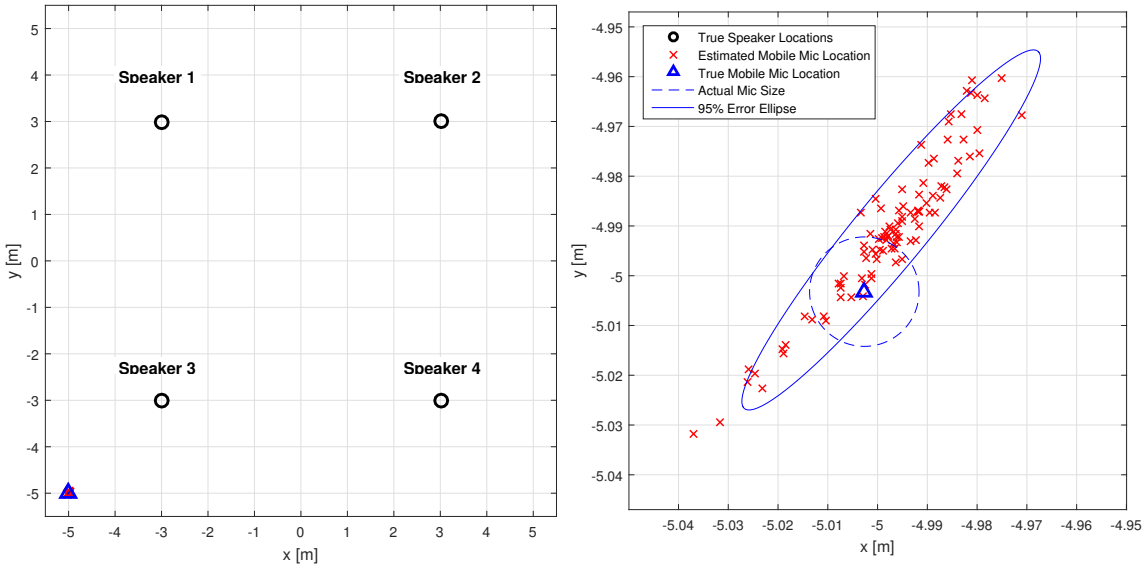


(a) Full test area **(b) Magnified at Testpoint 4**

Figure 28. Estimated location of mobile microphone at Testpoint 4

Compared to other Tier 2 testpoints, TDOA-based results for Testpoint 4 suffered in precision, as shown in Figure 28. The DOP of Testpoint 4 was 2.4, while previous testpoints were all less than 1.1, which helps explain the higher variance in the location estimates. While the orientation of the error ellipse was similar to the error ellipse for Testpoint for in Tier 1, the spread increased to make the error ellipse wider.

Testpoint 5.



(a) Full test area **(b) Magnified at Testpoint 5**

Figure 29. Estimated location of mobile microphone at Testpoint 5

Due to the high DOP, 7.15, at Testpoint 5, the variance of location estimates was much higher than the previous testpoints for Tier 2. As shown in Figure 29, 26 of the 100 estimates fell within the area of the microphone diaphragm. While the variance of location estimates was greater for Testpoint 5 than Testpoint 4, the bias was less, allowing for a similar DRMS of 2.34 cm.

3.6 Chapter Summary

Tier 1 used TOA measurements in an LSE to locate the mobile microphone from sound sources of known location with known signal structure and known timing emitting sequentially. Estimated drift correction was applied to compensate for receiver clock drift. Results showed cm-level accuracy, with the majority of estimates falling within the area of the microphone diaphragm.

Tier 2 used TDOA measurements in a similar LSE to locate the mobile microphone from sound sources of known location with known signal structure and unknown timing emitting sequentially. Because TDOA measurements cancel receiver clock error, drift correction was no longer necessary. Results were similar to those of Tier 1 with slight increases in the size of the error ellipses at several testpoints.

While the error in location estimation for each testpoint was relatively small compared to the size of the microphone diaphragm, there are several causes that may be attributed to the bias and variance of the estimates. Temperature measurements have a significant effect on the calculated speed of sound in Equation (8). For example, a variation of 1° Fahrenheit alters the speed of sound by approximately 0.3 m/s. While a calibrated thermocouple thermometer was used to measure temperature, the display only provided whole number Fahrenheit readings, which may cause bias due to the limited number of significant figures.

For Tier 1 experimentation, the clock drift correction in presented in Section 3.2.5 assumes a linear clock drift. The correction may have not been completely effective if the receiver clock drift had nonlinear effects. If the clock drift is nonlinear, the clock drift correction would still allow for bias in TOA measurements. Tier 2 and beyond implement a TDOA based solution which differences out clock error, making clock drift correction no longer necessary.

Noise generated in the test facility may lead to incorrect TOA estimates while de-

tecting peaks in the audio signals in Section 3.2.4. No noise sources were loud enough to generate a false peak, which would created a significant outlier compared to the other trials. However, it is plausible that noise in combination with the reverberant effects of the room and the playback of the speaker may have caused a false peak in the samples immediately following a true peak. A false peak of this kind would cause a TOA estimate slightly later than the true TOA, and result in a biased location estimate.

Finally, another cause of error is the calibration of the Vicon system to obtain truth data. The reflective spheres may not be exactly centered to the microphone and speaker diaphragms, which would cause the locations detected by the Vicon system to not be accurate. In addition, the Vicon system could have estimation errors and is subject to higher variance in its estimates for objects further from the center of the room due to DOP based effects.

IV. Positioning with Sequential Sound Sources at Unknown Locations

4.1 Chapter overview

In Chapter III, the solution required known locations of the sound sources which were acquired through the Vicon system in order to produce accurate location estimates of the mobile microphone. Tier 3 presents a more versatile system that does not require a priori knowledge of the sound source locations. TDOA measurements between the mobile microphone and each of the reference microphones were differenced to estimate not only the location of the mobile microphone, but also the locations of each sound source. By introducing six more reference microphones with known locations, each object was located with increased precision, and the DOP in the test area was reduced relative to previous tiers.

Before Tier 4, sound sources generated an impulse to allow for simplified peak detection for TDOA measurements. Building towards a more practical implementation, Tier 4 assumed the sound sources generate unknown coherent signals, such as human speech. Tier 4 implements the GCC method [17], described in Section 2.2 , to obtain TDOA measurements from more complex signal structures. Table 5 summarizes the conditions of testing for Tiers 3 and 4.

Section 4.2 presents the new methods to implement the additional reference microphones in the solution for Tier 3. Section 4.3 discusses the results of tests for Tier 3. Section 4.4 presents the new methods necessary to allow for TDOA measurements of unknown signals in Tier 4. Section 4.5 presents results for Tier 4 tests. Section 4.6 summarizes the methodology and results for Tiers 3 and 4.

Table 5. Conditions of testing for Tiers 3 and 4.

Tier	Sound Source Type	Sound Source Timing	Sound Source Location	Playback
3	Impulse	Unknown	Unknown	Successive
4	Recorded Speech	Unknown	Unknown	Successive

4.2 Tier 3 Methodology

4.2.1 Section Overview.

The methodology for Tier 3 is summarized in Figure 30. Methods updated for Tier 3 include initial location estimation through a closed form solution and TDOA based LSE of both sound source and speaker location. The Vicon system was still used to acquire truth data for comparison and evaluation purposes, but no longer provided sound source locations as a part of the solution. With the exception of methods discussed in this section, all testing was performed in the same manner as Tier 2.

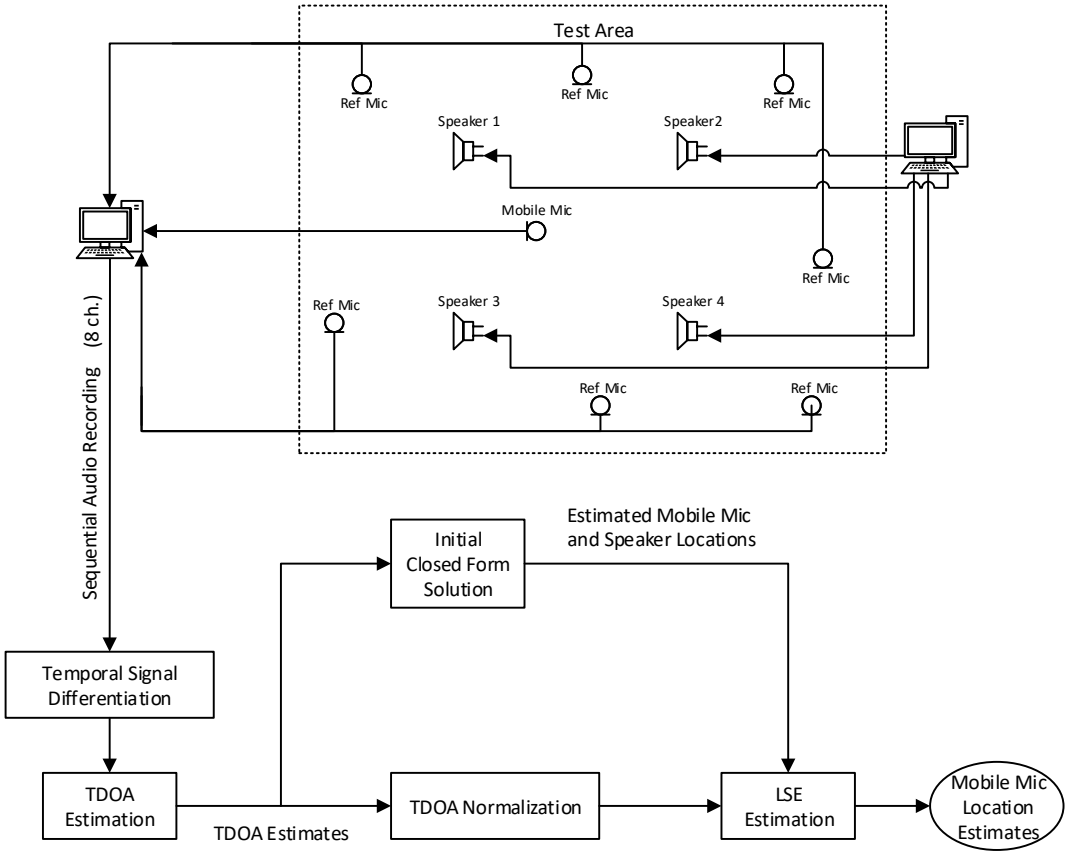


Figure 30. Methodology for obtaining mobile microphone location estimates in Tier 3.

4.2.2 Location Estimation Through Cascading Closed Form Solution.

The LSE estimation method in Tier 3 solves for not only the mobile microphone location, but also the unknown sound source locations. This more complex approach requires initial location estimates of the mobile microphone and each sound source. First, separate closed form solutions were used to estimate the location of each sound source based on the known locations of the reference microphones. The range differences of Equations (26 - 29) are between the locations of the reference microphones and one of the unknown sound sources. The x and y coordinates in Equations (27) and (28) are the coordinates of the reference microphones relative the the locations of

the first reference microphone. $\hat{\boldsymbol{x}}_{cf}$ was then assigned as the closed form estimate of the sound source location. This closed form solution was applied to each sound source. Then, a final closed form solution estimated the location of the mobile microphone using the newly estimated sound source coordinates in Equations (27) and (28). This closed form solution for the mobile microphone location is particularly susceptible to error; any error in the closed form solutions of the sound source locations cascade into the final mobile microphone closed form estimate. While increased error in location estimation is undesirable, the closed form estimates were only used as initial values for the LSE; the closed form estimates only have to produce general locations of the mobile microphone and sound sources to allow a convergent LSE result.

4.2.3 Least Squares Estimation for Speaker and Mobile Microphone Location.

The solutions of Tiers 1 and 2 required fixed, known locations of the sound sources. Because only the x and y coordinates and the timing of the receiver were unknown, the **H** matrix in Equation (20) required at least three linearly independent rows. In other words, estimating a single-point solution only requires at least three spatially separated sound sources with known locations to solve for the three unknown parameters. As Tier 3 does not assume a priori sound source locations are available, the LSE algorithm was changed to estimate the location of the sound sources while still producing an estimated location of the mobile microphone. In addition to requiring at least three sound sources, the solution for Tier 3 also requires at least three reference mobile microphones in order to estimate the location of the mobile microphone. TDOA measurements were made for each range difference between the mobile microphone and each reference microphone relative to each sound source. The range difference from the k^{th} sound source to the mobile microphone relative to the

l^{th} reference microphone is denoted as $\Delta\rho^{s_k}_{mr_l}$. Likewise, the vector of measurements for K sound sources and L reference microphones is

$$\boldsymbol{\Delta\rho}^{s}_{mr} = \begin{bmatrix} \Delta\rho^{s_1}_{mr_1}, & \Delta\rho^{s_1}_{mr_2}, \cdots, \Delta\rho^{s_1}_{mr_L}, & \Delta\rho^{s_2}_{mr_1} \cdots, \Delta\rho^{s_K}_{mr_L} \end{bmatrix}^T \tag{36}$$

.

Both K and L must be at least 3 in order to produce a single-point solution. The unknown state parameter for Tier 3 is

$$\boldsymbol{x} = \begin{bmatrix} x_m, & y_m, & x_{s_1}, & y_{s_1}, \ldots, x_{s_K}, & y_{s_K}, \delta t \end{bmatrix}^T. \tag{37}$$

where x_m and y_m denote the coordinates of the mobile mic, x_{s_1} and y_{s_1} through x_{s_K} and y_{s_K} denote the location of each sound source, and δt denotes the receiver clock error common to all microphones. Likewise, the initial state vector for the LSE equation includes the initial estimates of the sound source locations:

$$\boldsymbol{x}_0 = \begin{bmatrix} x_{m_0}, & y_{m_0}, & x_{s_{1_0}}, & y_{s_{1_0}}, \ldots, x_{s_{K_0}}, & y_{s_{K_0}}, \delta t_0 \end{bmatrix}^T. \tag{38}$$

The $\mathbf{H}$ matrix is then constructed according to Equation (18) with the measurement, unknown state, and initial state vectors from Equations (36 - 38). The resultant LSE solution, $\hat{\boldsymbol{x}}$ from Equation (21), includes the estimated coordinates of the mobile microphone and each sound source. Although estimating the location of mobile microphone and time offset was the main objective, the sound source location estimates may be a useful byproduct in application.

4.3 Tier 3 Results

4.3.1 Comparison of Dilution of Precision.

In previous tiers, the DOP was only dependent on the configuration of the mobile microphone and sound sources with known location. In Tier 3, the solution estimates the location of the mobile microphone and each sound source. Because the solution uses additional reference microphones, the location of those reference microphones also affects the DOP of Tier 3 results. Table 6 includes DOP values at each testpoint for Tiers 3-5. Figure 31 is a DOP map of the test area for Tier 3. Because Tiers 4 and 5 use the same configuration of sound sources and reference microphones and estimate the same quantities, Figure 31 may also be referenced for the related tests. While the sound sources are at unknown locations, the sound sources are kept at fixed locations for consistent testing. If the sound sources were to be relocated, DOP values may change. Note that while DOP values less than one are uncommon in other applications such as GPS, sub-unity DOP is possible [28]. GPS solutions are typically limited to using one receiver and four satellites, whereas this configuration utilizes all seven reference microphones and four sound sources.

4.3.2 Tier 3 Results.

Table 6 gives a summary of results for Tier 3 at each testpoint. Because of the bias in measurements, a direct comparison between DOP and DRMS was difficult to observe. DOP predicts higher variance in non-biased estimates at locations with a given speaker and microphone configuration whereas DRMS accounts for both variance and bias. For Tiers 3-5, only select testpoints of interest with uncommon results are discussed in depth.

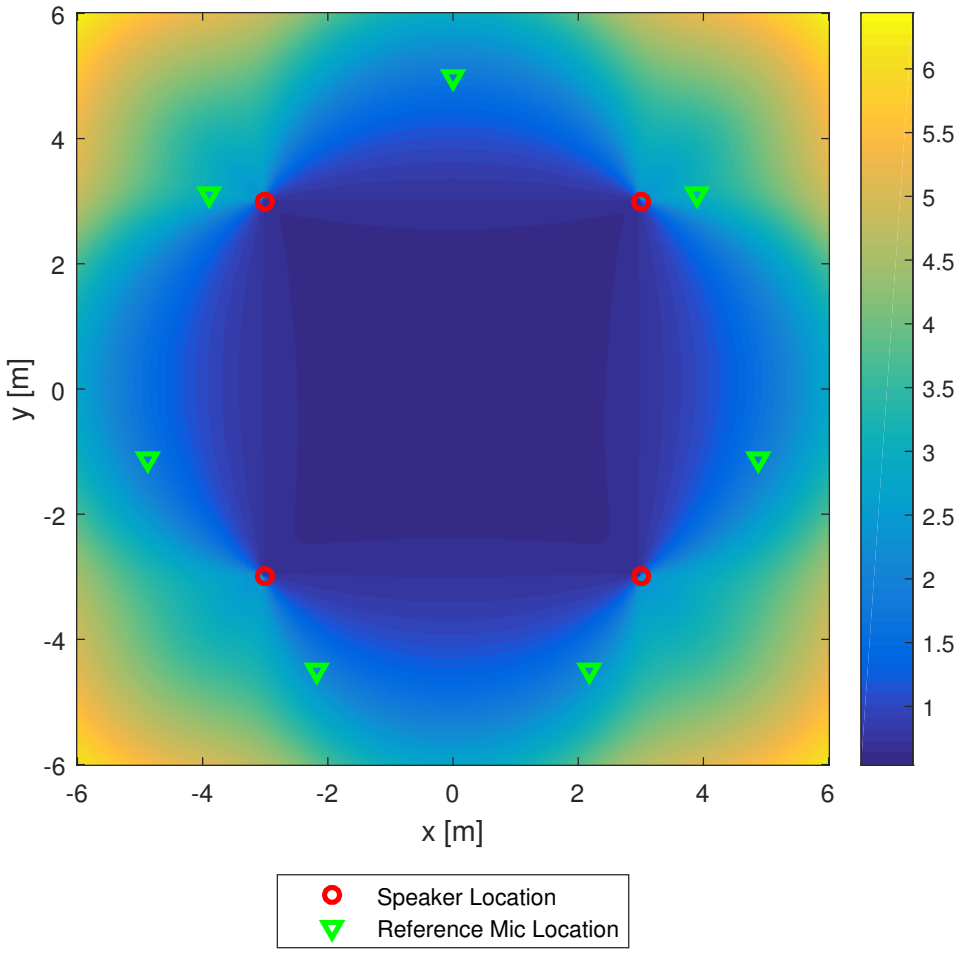
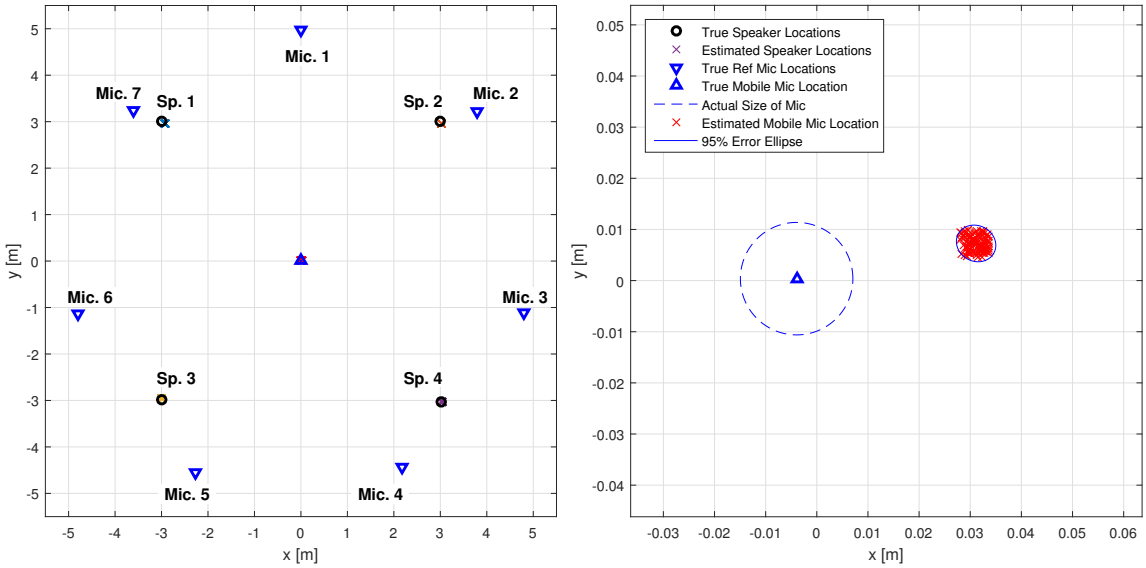


Figure 31. DOP map of testbed for Tier 3, 4, and 5 tests. Color corresponds to DOP as a function of the location of the mobile microphone.

Table 6. Results for Test Performed in Tier 3

Testpoint	Approx. Coords.	DOP	Clock Error (μs)		Position Error (cm)				
			δt	$\sigma_{\delta t}$	$\bar{x}$	σ_x	$\bar{y}$	σ_y	DRMS
0	(0,0)	0.5374	-13.68	0.76	3.51	0.15	0.69	0.15	3.58
1	(1,-2)	0.5938	-16.79	1.34	1.63	0.16	-0.76	0.17	1.65
2	(-2,-2)	0.5990	-78.12	0.67	-2.69	0.15	-1.19	0.16	2.95
3	(2.9,2.9)	0.6437	-84.32	0.64	-2.43	0.15	-1.45	0.16	2.24
4	(4,-2)	1.447	67.48	29.77	-0.98	0.25	0.89	0.17	1.24
5	(-5,-5)	4.554	-205.34	17.34	-6.03	0.48	-1.08	0.46	6.16

Testpoint 0.

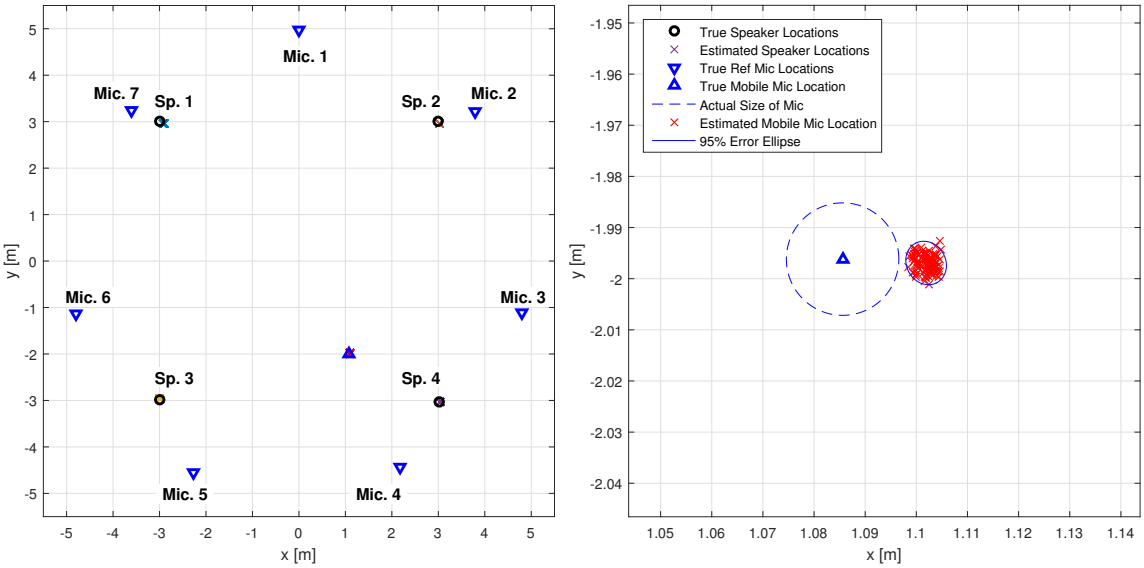


(a) Full test area **(b) Magnified at Testpoint 0**

Figure 32. Estimated location of mobile microphone at Testpoint 0

The test area configuration and results for Testpoint 0 are shown in Figure 32. Despite the more complicated process of locating the mobile microphone using sound sources with unknown locations, the estimates still fell within the general area of the true locations. The bias of the estimations brought all estimates outside of the area of the microphone diaphragm. As expected with the reduced DOP from using multiple reference microphones, the size of the error ellipse was much less than previous tiers as shown in Figure 32b. The DRMS was significantly higher than previous tests at 3.58 cm, which is mostly due to the bias of the estimates rather than the variance. This bias may be explained by the LSE estimating not only the location of the mobile microphone, but also the four sound sources. The LSE maximizes the likelihood of estimating $\boldsymbol{x}$ exactly, with making no guarantees of minimizing bias. While not the primary objects of interest, the sound sources were located reasonably well with all estimates within 5 cm of the respective speaker.

Testpoint 1.

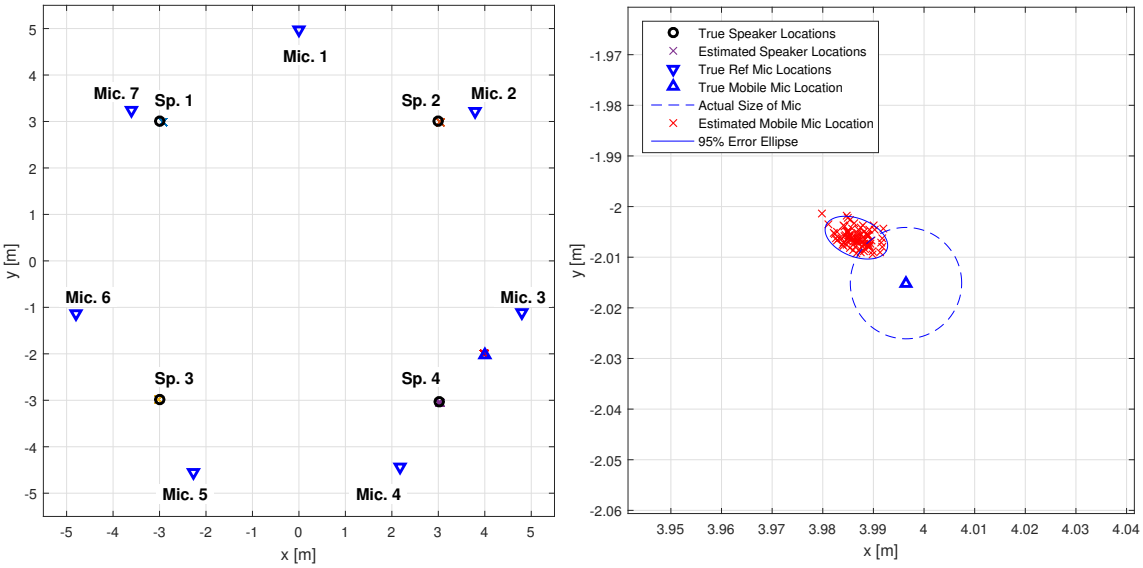


(a) Full test area **(b) Magnified at Testpoint 1**

Figure 33. Estimated location of mobile microphone at Testpoint 1

The test area configuration and results for Testpoint 1 are shown in Figure 33. The bias of estimates in Tier 3 for Testpoint 1 was greater than that of Tier 2. All estimates fall outside the microphone diaphragm. However, the variance and thus the error ellipse of the estimates was much smaller because of the reduced DOP from the new configuration involving 7 reference microphones. The DRMS of the estimates was 1.65 cm, increasing by 6 mm compared to Tier 2. This increase was due to the bias of the estimates despite the increased precision.

Testpoint 4.

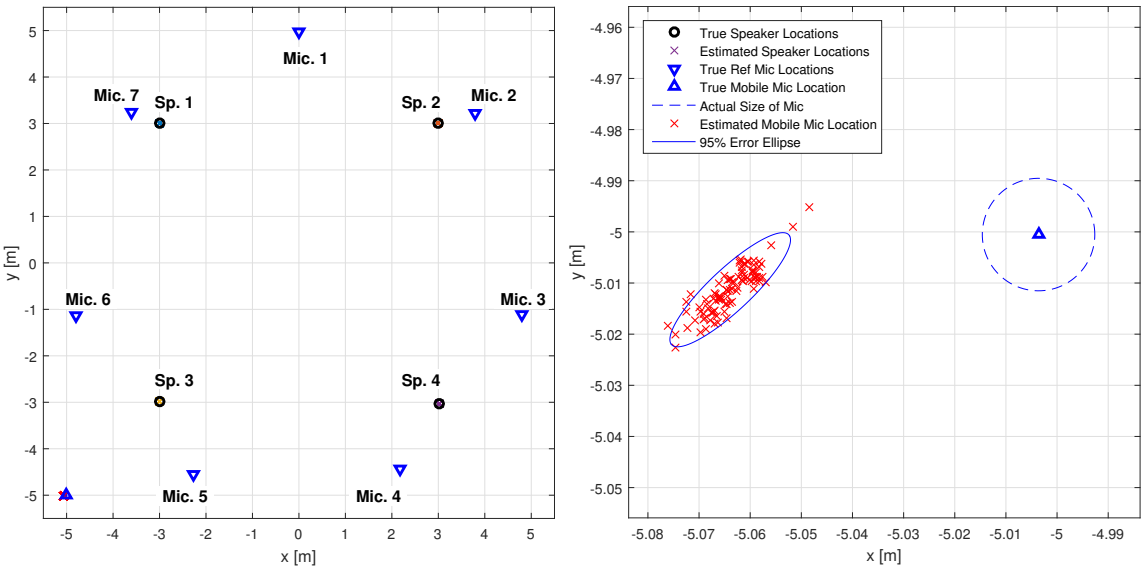


(a) Full test area **(b) Magnified at Testpoint 4**

Figure 34. Estimated location of mobile microphone at Testpoint 4

The test area configuration and results for Testpoint 4 are shown in Figure 34. In previous tiers, Testpoint 4 showed significant increase in the size of the error ellipse due to increased DOP. With the new configuration introduced in Tier 3, the DOP of Testpoint 4 is reduced from 2.4 to 1.4. As expected with the decrease in DOP, the size of the error ellipse for the testpoint in Tier 3 also diminished as shown in Figure 34b. The bias of the estimates is comparable to the preceding testpoints for Tier 3 with 15 of the estimates falling within the area of the microphone diaphragm. The DRMS of the location estimates was 1.24 cm, which was also less than that of Tier 2.

Testpoint 5.



(a) Full test area **(b) Magnified at Testpoint 5**

Figure 35. Estimated location of mobile microphone at Testpoint 5

The test area configuration and results for Testpoint 5 are shown in Figure 35. While the orientation of the error ellipse for Testpoint 5 remained the same between Tier 3 and previous tiers, the size of the ellipse was much smaller, showing more consistent estimates. However, a stronger bias is present. Estimates fell left of and slightly below the true microphone location, similar in magnitude to the bias found at Testpoint 0. The DRMS of Testpoint 3 was nearly triple that of Tier 2 due to the increased bias in the estimates.

4.4 Tier 4 Methodology

4.4.1 Section Overview.

Tiers 1-3 all assumed the signals emitted from the sound sources were impulses. A time domain amplitude peak detector was sufficient in accurately determining TOA and TDOA measurements. Tier 4 does not assume sound sources produce impulse waveforms. Instead, signals consist of recorded human speech. Because the signals are more complex waveforms not guaranteed to have consistently timed peaks, direct peak detection is not sufficient in determining TDOA measurements and instead, a GCC method is implemented [17], as described in Section 2.2. Changes in the methodology discussed in this section include the standard used to collect and differentiate between the recorded human speech sound sources, as well as the GCC method of collecting TDOA measurements.

4.4.2 Signal Collection and Differentiation.

The method for collecting audio for Tier 4 was similar to the approach used in previous tiers. As before, each speaker successively emitted one after the other, but the recording time for each signal was increased to two seconds, with a half second pause between each speaker. An increased timeframe for each speaker allowed a higher probability of a strong peak with the GCC TDOA method. The half-second pause between each speaker prevents interference via reverberation overlapping between trials. The four sequentially played sounds with half second pause are evident in the waveform recorded from each microphone, plotted in Figure 36. Because each trial takes much longer to perform, 20 trials were taken instead of 100.

4.4.3 Generalized Cross-Correlation Time Difference of Arrival Measurements.

The amplitudes of the recorded signals vary depending on the distance from each microphone to the emitting sound source. But, the general shape of the waveform is maintained across the recordings from the spatially separated microphones, as seen in Figure 36. Instead of direct peak detection, the signal received from the mobile microphone is cross-correlated with each of the signals from the reference microphones according the the GCC method outlined in Section 2.2 [17]. Because the general shapes of the waveforms are similar with some time offset, a peak forms at the point of the time-offset of the microphones. The time offset of the peak in the cross correlation corresponds to the TDOA measurement between the mobile microphone and reference microphone for that sound source.

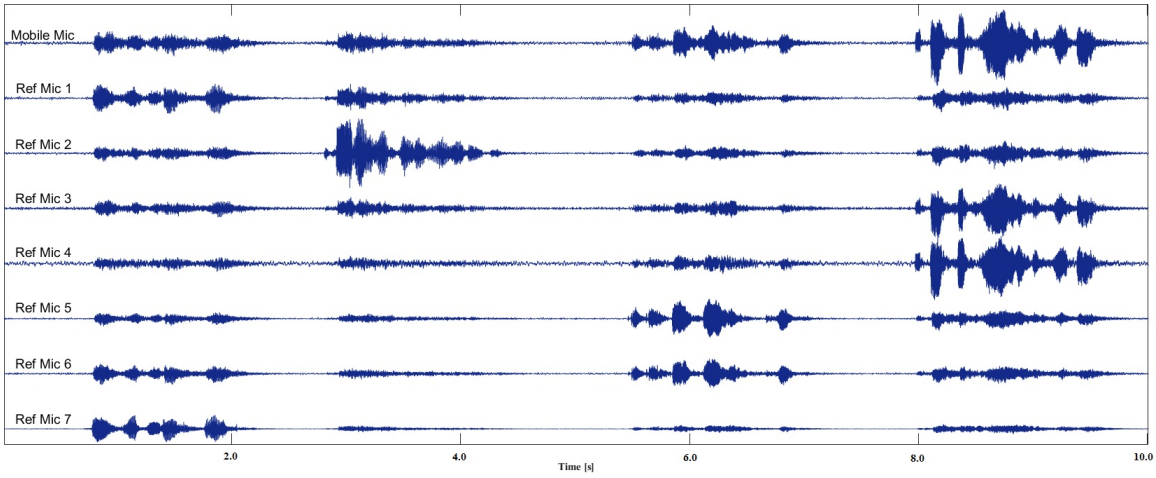


Figure 36. Example of multi-channel audio recording for one trial.

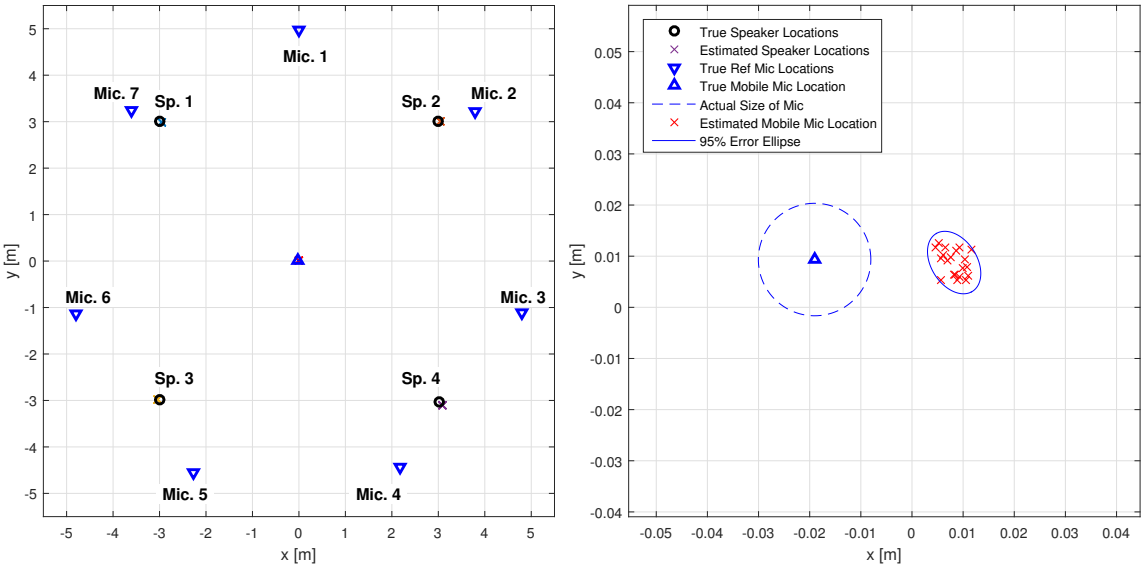
4.5 Tier 4 Results

Table 7. Results for Test Performed in Tier 4

Testpoint	Approx. Coords.	DOP	Clock Error (μs)		Position Error (cm)				
			$\bar{\delta t}$	$\sigma_{\delta t}$	$\bar{x}$	σ_x	$\bar{y}$	σ_y	DRMS
0	(0,0)	0.5374	-90.88	0.67	2.73	0.21	-0.06	0.25	2.75
1	(1,-2)	0.5938	-25.69	0.91	1.98	0.17	-0.72	0.18	2.12
2	(-2,-2)	0.5990	-6.27	3.24	2.00	0.15	1.05	0.15	2.02
3	(2.9,2.9)	0.6437	-13.62	0.53	4.30	0.18	-0.80	0.17	4.38
4	(4,-2)	1.447	38.39	19.70	-1.26	0.27	1.71	0.19	2.15
5	(-5,-5)	4.554	120.04	20.96	-6.04	0.57	-1.75	0.66	6.34

As with Tier 3, because DRMS values are more weighted by bias than variance, direct comparison between DOP and DRMS is difficult, with the exception of Testpoint 5, where a large increase in DOP is concurrent with a relatively large DRMS value.

Testpoint 0.

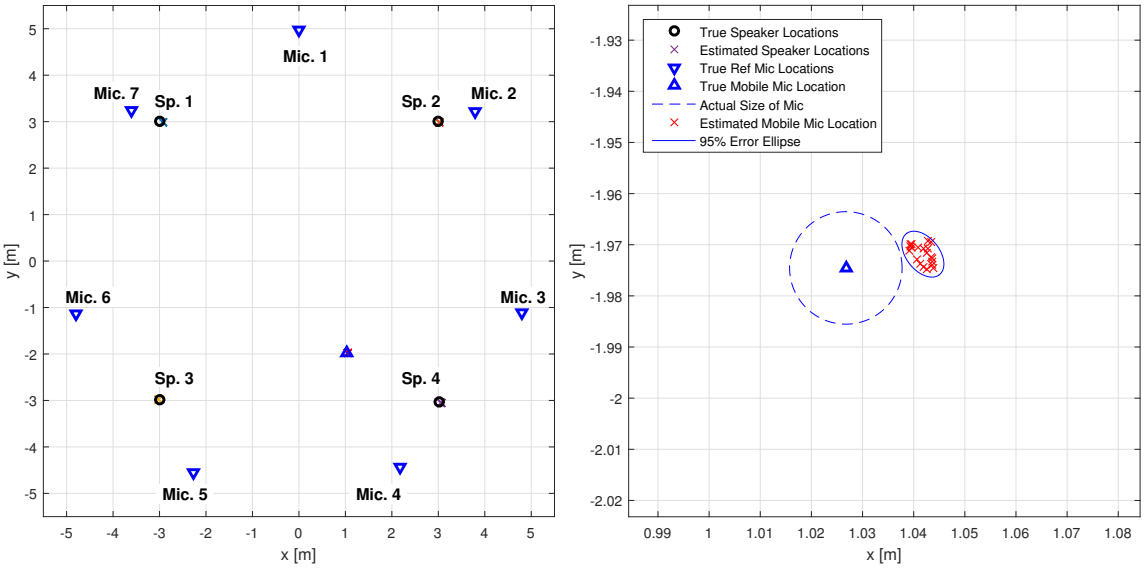


(a) Full test area **(b) Magnified at Testpoint 0**

Figure 37. Estimated location of mobile microphone at Testpoint 0

The bias of the estimates approximately 1.5 cm outside the area of the microphone diaphragm, as shown in Figure 37. Results were fairly consistent with those of Tier 3 for Testpoint 0. The consistency of bias shows that using GCC method for obtaining TDOA as opposed to amplitude peak detection did not have a significant impact on the accuracy of the location estimates when sound sources were not close to the microphone.

Testpoint 1.

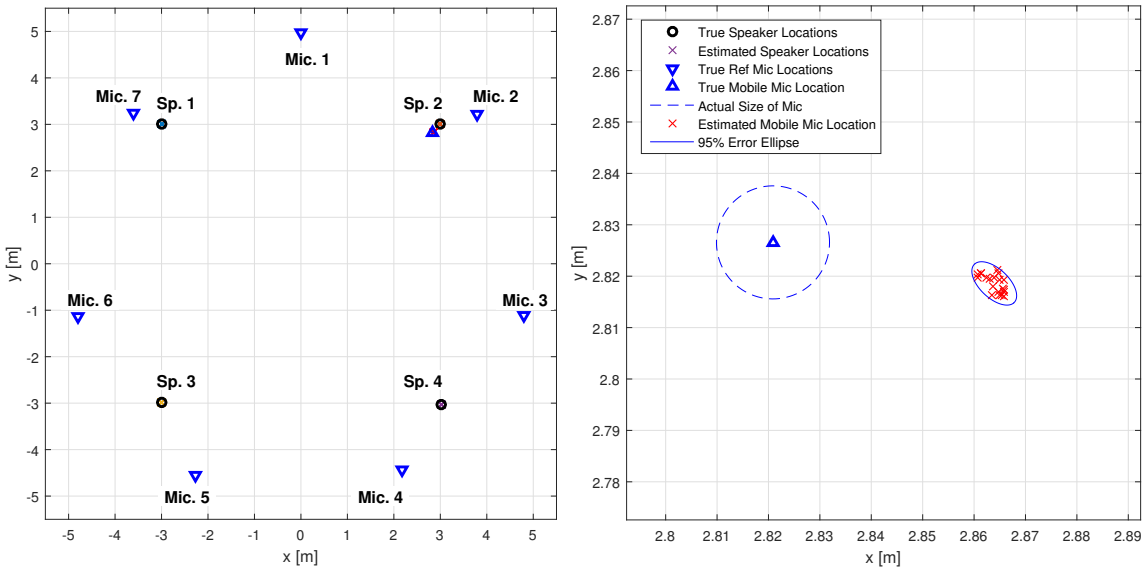


(a) Full test area **(b) Magnified at Testpoint 1**

Figure 38. Estimated location of mobile microphone at Testpoint 1

The test area configuration and results for Testpoint 1 are shown in Figure 38. At Testpoint 1, the biases present in Tier 4 testing were again similar to the biases present in Tier 3, giving further evidence to support that the GCC method has minimal effect on the accuracy of the location estimation when sound sources are not directly near the microphone.

Testpoint 3.



(a) Full test area **(b) Magnified at Testpoint 3**

Figure 39. Estimated location of mobile microphone at Testpoint 3

Tier 4 showed a significant increase in DRMS for Testpoint 3, which was mostly weighted by the bias of the estimates, shown in 39. The difference between Tiers 3 and 4 is the signal structure (impulse vs. speech) and the method of TDOA acquisition (amplitude peak detection vs. GCC method). It is likely that the new method of acquiring TDOA measurements decreases in accuracy when microphones are adjacent to any of the sound sources.

4.6 Chapter Summary

Tier 3 used TDOA measurements in an LSE to locate the mobile microphone from sound sources of unknown location with known signal structure and unknown timing emitting sequentially. In order to do so, seven reference microphones were introduced. As a bi-product, the LSE estimated the general locations of the sound sources in addition to the location of the mobile microphone. Results showed slightly degraded accuracy, yet increased precision compared to Tier 2.

Tier 4 used TDOA measurements in a similar LSE to locate the mobile microphone from sound sources of unknown location with unknown signal structure and unknown timing emitting sequentially. In order to acquire TDOA measurements, the GCC method was implemented. Results were similar to those of Tier 3 at testpoints not adjacent to sound sources. Testpoint 3, close to a sound source, showed decreased accuracy in location estimation for Tier 4.

V. Positioning with Simultaneously Emitting Sound Sources at Unknown Locations

5.1 Chapter Overview

In previous tiers, sound sources were played sequentially so that the audio from any speaker would not interfere with the location estimation of another speaker. Tier 5 allowed for more a more realistic application where sound sources emitted simultaneously, as outlined in Table 8. Section 5.2 presents the methodology for Tier 5, followed by results in Section 5.3, and conclusion in Section 5.4.

5.2 Tier 5 Methodology

5.2.1 Section Overview.

This section introduces Blind Source Separation (BSS) as a means of separating the sound sources in the test area to recreate each signals as originally emitted without the interference of the other signals. In order to separate the signals, the technique for separating the signals requires known or estimated locations of the sound sources. Because the sound source locations are unknown, SRP mapping [13], first discussed in Section 2.4.4, was used to estimate the locations. While SRP mapping is successful in detecting sound sources, the maps often contain numerous false positives. Peak Isolation Filtering (PIF) was used to dramatically reduce the number of false positives with a high likelihood of maintaining true peaks. With the estimated locations of sound sources, TFM was used to isolate the signals of each sound source from one another. Because SRP mapping with PIF did not always accurately estimate the sound source locations, a method was introduced for eliminating errant sound source estimations before affecting the LSE of the mobile microphone. Figure 40 is a schematic showing the order in which these methods are implemented in Tier 5.

Table 8. Conditions of testing for Tier 5.

Tier	Sound Source Type	Sound Source Timing	Sound Source Location	Playback
5	Recorded Speech	Unknown	Unknown	Simultaneous

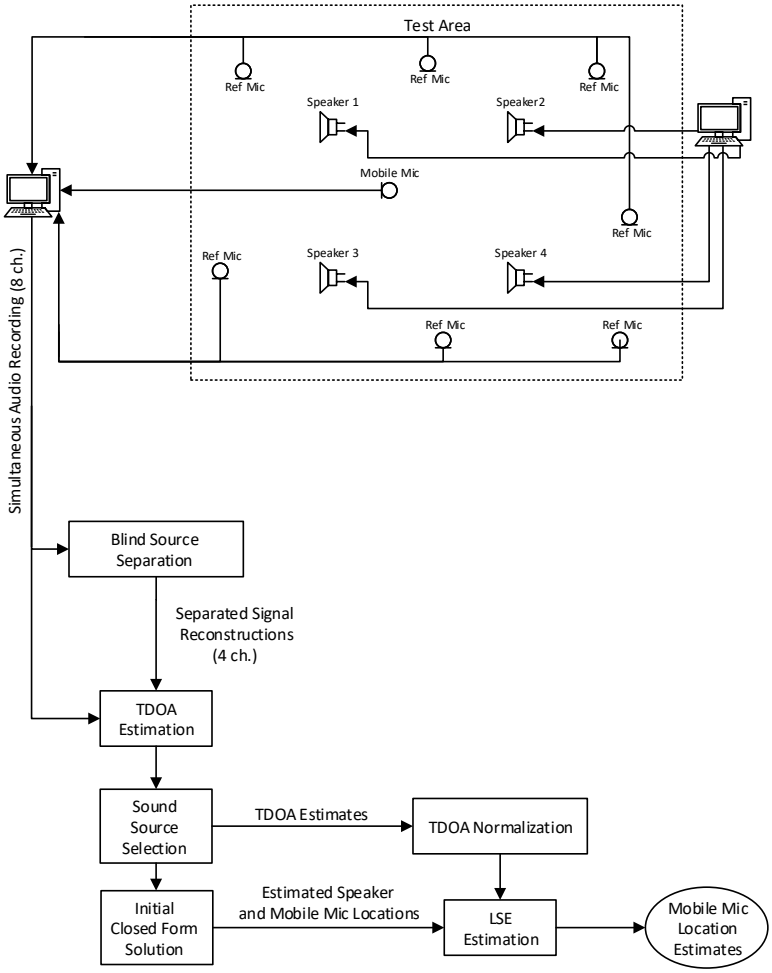


Figure 40. Methodology for obtaining mobile microphone location estimates in Tier 5.

5.2.2 Time Frequency Masking.

In Tier 5, sound sources simultaneously emit. If microphone output were directly correlated with other microphone output, as done in Tier 4, there would be multiple peaks in the correlation caused by the four sound sources, with no indication of which

peak corresponds to a certain sound source. In order to avoid ambiguity attribution of correlation peaks to sound sources, the signals recorded from the microphones were reconstructed to isolate the audio from each of the sound sources. Isolated recordings of each of the signals were then correlated with the recordings from each of the microphones in order to determine distance differences.

The signals were isolated through the process of TFM. First, the original recording were beamformed; each channel was proportionally delayed according to the distance between the known microphone locations and estimated sound source locations obtained through SRP (presented in Section 5.2.3). The delayed channels were then summed together so that the signal of interest constructively added across the channels. Overlapping Short Time Fourier Transforms were then calculated for each of the beamformed signals to create a Time-Frequency (TF) map. Power differentials between frequency regions in the map indicate if a given area of the map may be associated with the signal of interest. Areas that were not associated with the Signal of Interest (SOI) were then set to 0 in a binary mask. The mask is then multiplied by the TF map and inverse transformed to create the isolated reconstruction of the signal of interest [30]. Figure 41 outlines the process of TFM to generate the reconstructed signal. TFM was performed with each of the sound sources as the SOI to create reconstructions of each original sound source. Depending on the relative power of the sound sources and the accuracy of the SRP sound source location estimates obtained in Section 5.2.3, some signal reconstructions may not be strong enough to isolate the signal. Section 5.2.6 discusses methods to determine if the reconstructed signals sufficiently isolate the signal of interest for use in locating the mobile microphone.

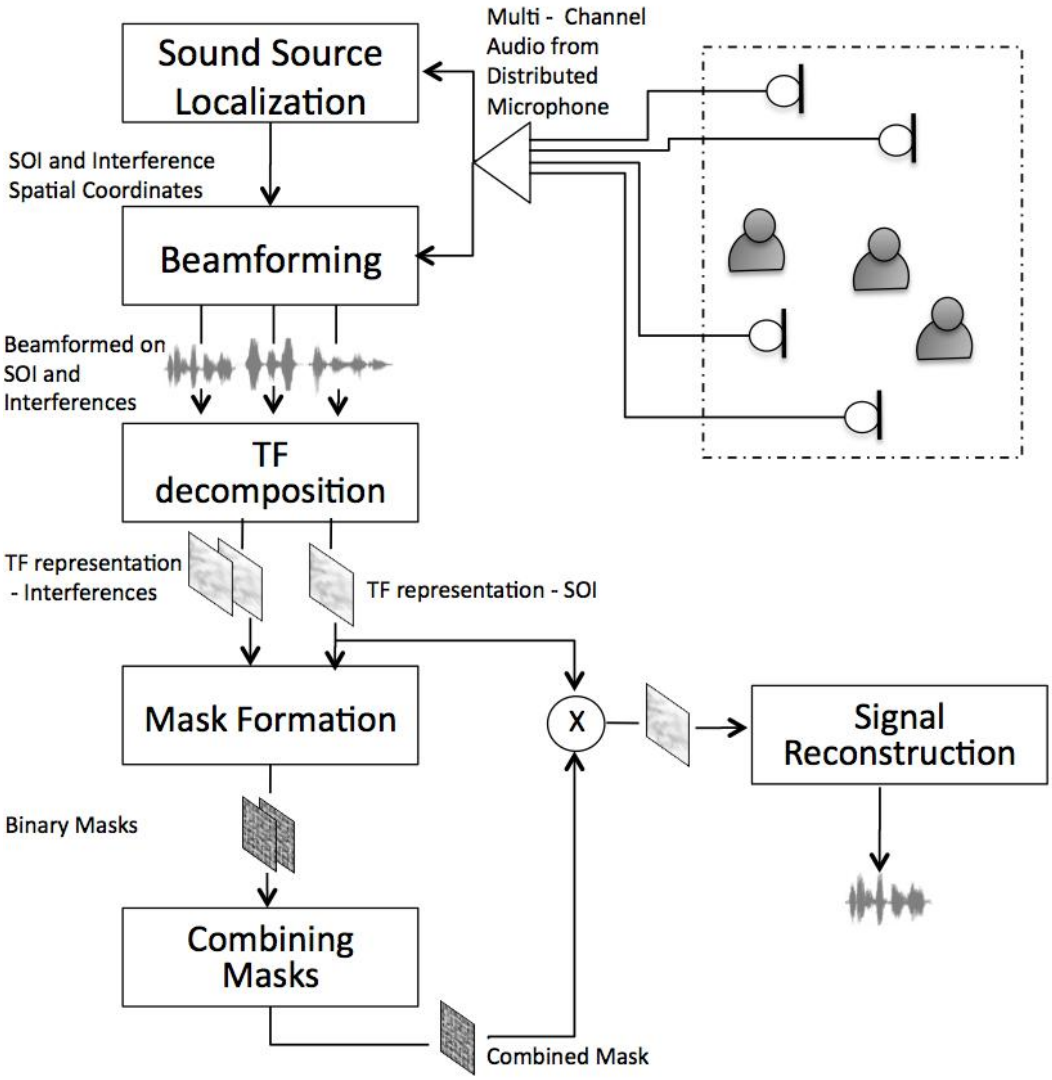


Figure 41. TFM speaker of interest extraction system. Adapted from [30].

5.2.3 Steered Response Power.

In order to perform the beamforming in TFM, either estimated or known locations of sound sources are required. Tier 3 and beyond assume unknown locations of the sound sources, so the locations must be estimated prior to LSE estimation. SRP allows an estimation of the sound source locations only using the audio recorded from the known reference microphone locations. The SRP technique was used to map the

test area with the microphone channels beamformed for each pixel in the map. The power of the delayed and summed signal was recorded at each point in the map. In areas where sound sources were active, the microphone channels constructively added, resulting in higher signal power than areas where the signals destructively combined. Areas in the map with high power indicate the probable location of a sound source.

If performed at the original 44.1 kHz sampling frequency of the recorded audio, SRP becomes a computationally burdensome process. The pixel resolution of the SRP image must be strong enough to not skip over any samples in the audio between pixel locations. With too weak of a resolution, the center of the pixels may not fall on the location of the sound sources, so sound sources can be missed. As the resolution of the map increases by N, the number of computations required to create the map increases by N^2. At 44.1 kHz, approximately a 7 mm pixel resolution is required, translating to a map area of 420,000 delay and sum computations. In order to reduce the computation required, the recorded audio is filtered and downsampled to 8 kHz. Filtering and downsampling allows for larger pixel size with a much lower chance of skipping over the highest powered alignment of the delayed and summed signal. At 8 kHz, approximately a 4 cm pixel is required, which translates to only 600 delay and sum operations required to create the SRP map.

With multiple simultaneously emitting sound sources, one or two speakers may dominate other sound sources. However, there may be instances of time where the dominate speakers are silent, allowing for a clearer location estimate of softer speakers. The SRP algorithm was calculated for multiple time windows during the trial and the power was averaged across the time windows to allow for more accurate location of multiple sound sources. Figure 42 shows a time averaged SRP map revealing the general locations of four sound sources.

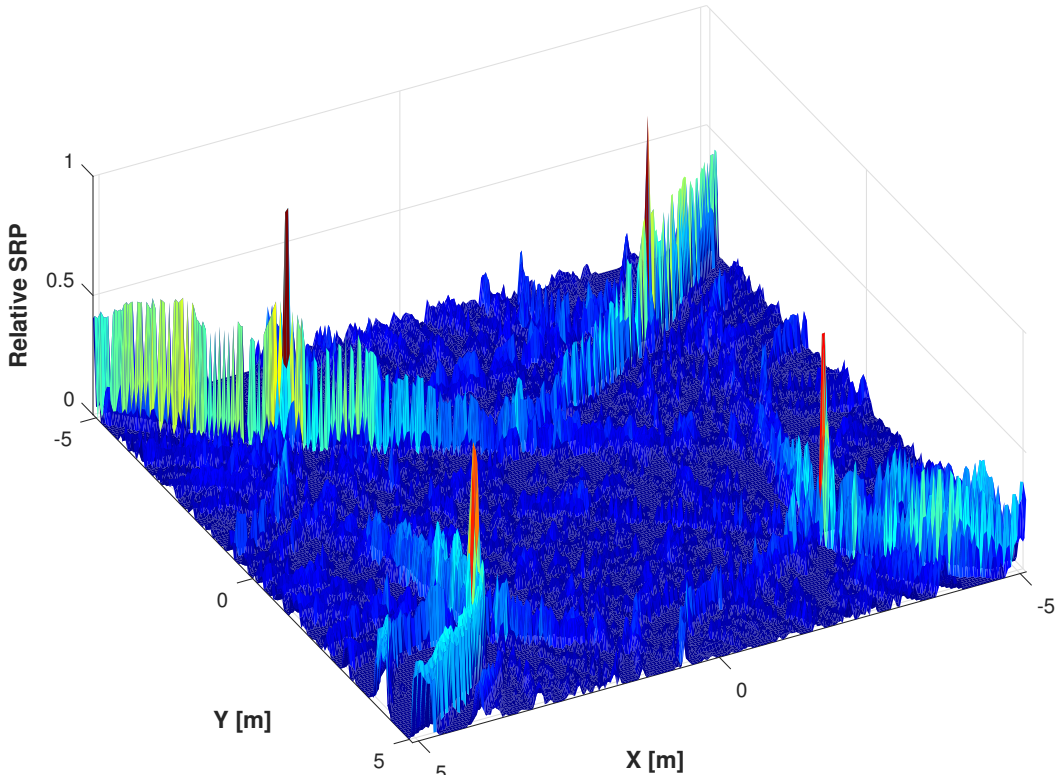


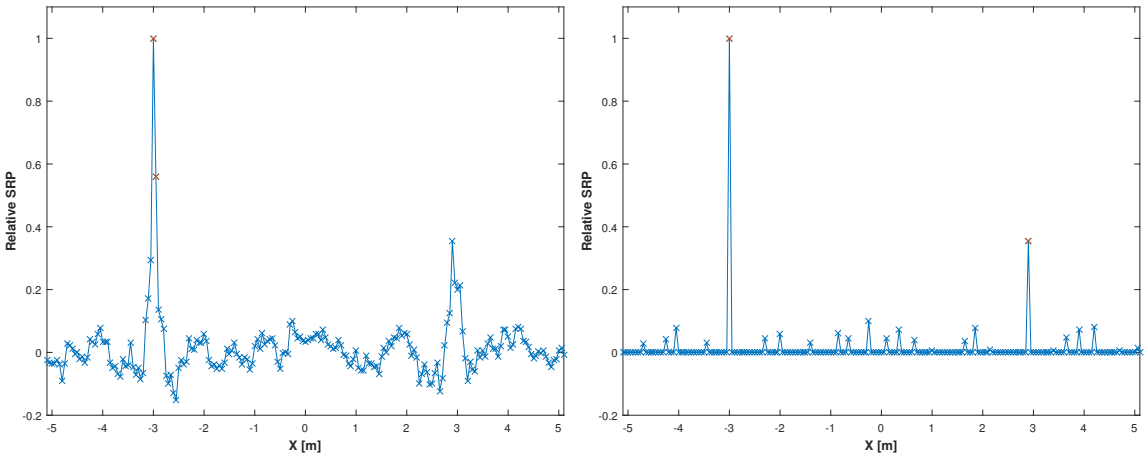
Figure 42. Time averaged SRP map to showing all four sound source locations.

5.2.4 Peak Isolation Filtering.

There are two main issues with resolving the speaker locations solely using SRP: streaking and multiple high power valued pixels near a single sound source. Streaking is present in locations where the signal power is predominately generated by only two microphone recordings. When three or more microphones significantly contribute to the SRP, a larger peak is apparent. However, the streaking caused by one sound source may generate more power than a single peak generated by another. Or, a sound source may have pixels adjacent to the true location that are greater than the highest peak of another sound source. The four largest power values on the map may not correspond to the four sound sources; pixels near higher powered sound sources may have greater power than the highest peak of softer sound sources and cause errant estimates of the sound source locations.

PIF was used to produce a more defined center of each of the peaks and to reduce the effects of streaking. Pixels in the SRP map were saved only if they were the highest value pixel of all their nearest neighbors. Otherwise, the pixels are set to 0, effectively creating a mask around the highest values in the map. This procedure assumes that no two sound sources were located close enough to both fall under the area of the mask. A larger mask produces fewer false positive peaks, but risks masking one sound source if it is too close to another sound source. A smaller mask is more likely to maintain all sound sources, at the expense of keeping more false positive peaks. For the results shown, a mask with an area of 7 x 7 pixels was used which assumes no two sound sources are within three pixels (i.e. 12 cm) of one another.

Figure 43a shows a simplified example of peak isolation filtering along one dimension. Figure 43a is a cross-section of Figure 42 at $Y = 3$ m. Without peak isolation filtering, two sound sources are detected around $X = -3$ m, since the two highest values, shown in red, are adjacent to one another. With the application of peak isolation filtering, as shown in Figure 43b, both sound sources are detected: one near $X = -3$ m and the other near $X = 3$ m.



(a) **(b)**

Figure 43. (a) Cross-section of unfiltered SRP map shown in Figure 42 at Y = 3 m. Estimated sound source locations (red) are both on a single peak. (b) SRP cross-section after PIF has been applied, producing correct sound source location estimates.

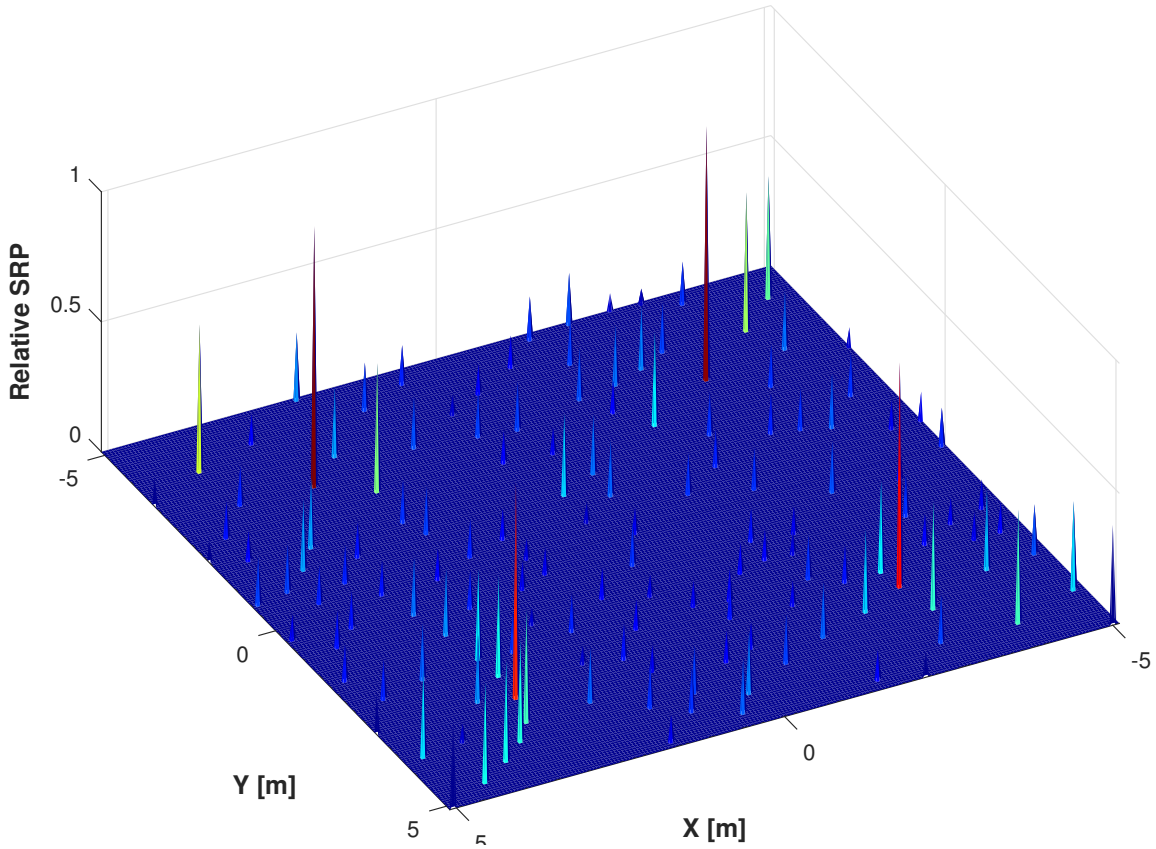


Figure 44. SRP map from Figure 42 with peak isolation filtering applied. Streaking has been reduced and four distinct peaks are shown corresponding to the source locations.

Figure 44 shows the time averaged SRP map in Figure 42 after peak isolation filtering was applied to the SRP map. While some points along the original streaks were still apparent after applying peak isolation filtering, the larger values along the streaks that were most likely to cause false estimations were eliminated. Assuming four sounds sources within the test area, the coordinates of the four remaining pixels with the highest value were used as the estimated locations of the sound sources. The location estimates of the sound sources were then implemented into the TF masking algorithm in order to produce the reconstructed signals necessary for determining TDOA measurements.

5.2.5 Time Difference of Arrival Measurements.

Because the sound sources were simultaneously emitted in Tier 5, a new approach to obtaining the TDOA measurements was necessary. If the same GCC method used in Tier 4 were implemented, there would be multiple peaks in the correlation function with no clear indication of which peak corresponds to a particular sound source. Instead, two separate GCCs are performed: one between the audio from the mobile microphone and the reconstructed audio of a given sound source, and the other between the audio of a reference microphone and the same reconstructed audio, as shown in Figure 45. Because the reconstructed audio carried less noise from other sound sources, the correlation functions were more likely to result in single peaks. The TDOA measurement relative to the mobile microphone and the sound source of interest is the difference in timing of the two correlation function peaks.

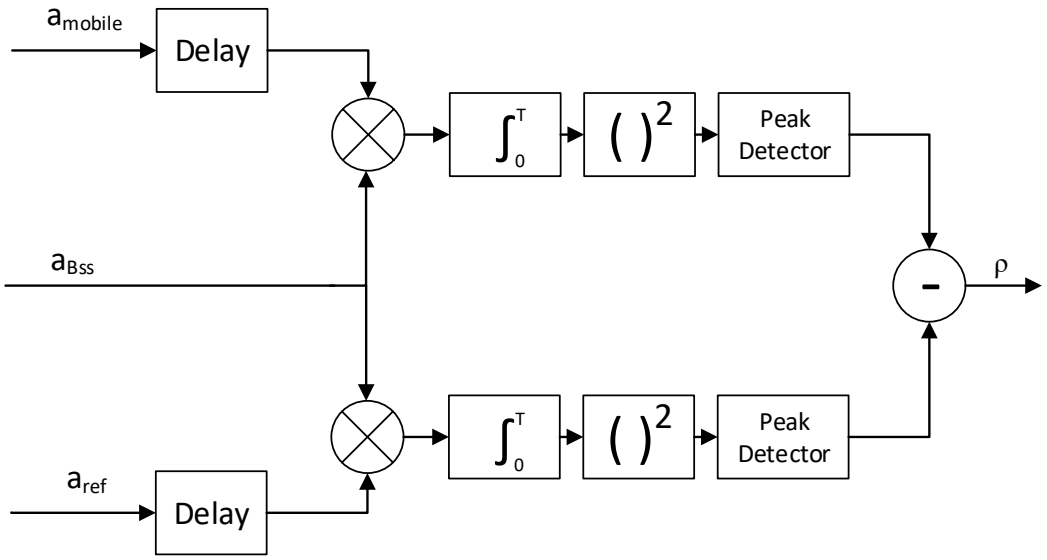


Figure 45. Modified GCC method, applying reconstructed audio for generating stronger, unambiguous correlation peaks. Peaks of the mobile microphone correlation and the reference microphone correlation are then differenced to determine the TDOA value.

5.2.6 Sound Source Selection.

The new approach to obtaining the TDOA measurements reduces interference from other sound sources. But, certain conditions may still cause false peaks in the correlation functions, leading to errant TDOA measurements. For example, if one of the sound sources is significantly louder than the other sound sources, the reconstructed audio of a weaker sound source may still contain significant traces of the louder sound source and cause a peak corresponding to the location of the louder sound source. As stated in Section 4.2.3, only three sound sources and three reference microphones are necessary to estimate the location of the mobile microphone. In testing, four sound sources were used, so that if one sound source caused multiple ambiguous peaks in the cross correlation function, the TDOA measurements from that speaker could be discarded, and the LSE estimate could be obtained with the remaining three sound sources. In testing, if a cross correlation function produced a peak with at least twice the amplitude of all other points, the produced peak was determined to be sufficient for use in the TDOA estimation. However, if no significantly dominant peak was produced, then the TDOA measurements corresponding to that sound source were discarded. Figure 46a gives an example where both cross correlation functions produced clear dominant peaks, whereas Figure 46b shows cross correlation functions with multiple, ambiguous peaks, indicating the TDOA measurement for that sound source should be discarded.

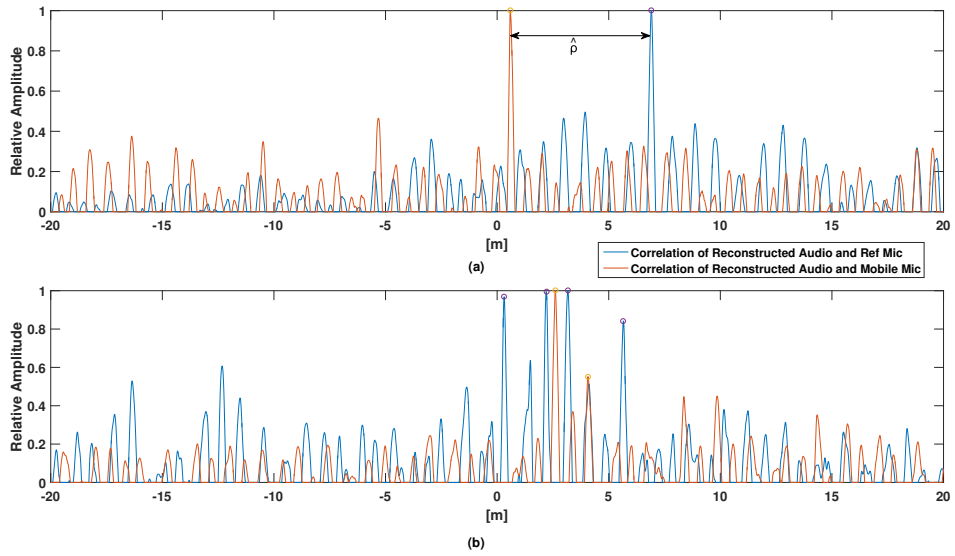


Figure 46. (a) Cross-correlation example with good peak determination. (b) Cross-correlation example with ambiguous peak determination.

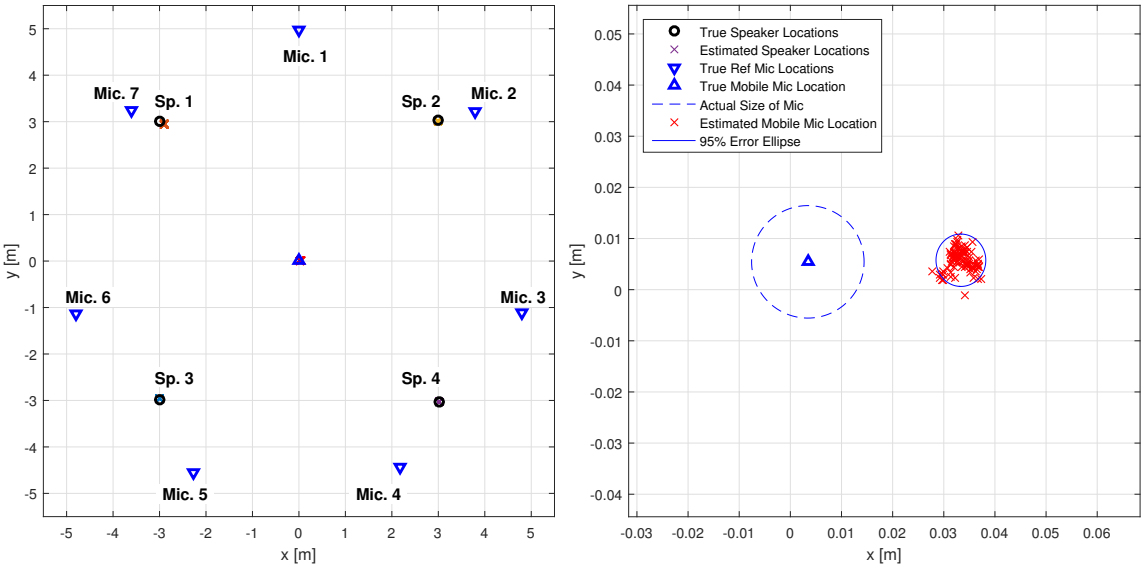
5.3 Tier 5 Results

Table 9 provides a summary of results for Tier 5 at all testpoints. While most testpoints showed results similar to those of Tier 4, Testpoint 3 showed a significant increase in DRMS, far more than any other test performed. Testpoint 3 also showed the highest clock error at 327.29 μs.

Table 9. Results for Test Performed in Tier 5

Testpoint	Approx. Coords.	DOP	Clock Error (μs)		Position Error (cm)				
			$\bar{\delta t}$	$\sigma_{\delta t}$	$\bar{x}$	σ_x	$\bar{y}$	σ_y	DRMS
0	(0,0)	0.5374	57.19	5.04	2.99	0.19	0.03	0.17	3.00
1	(1,-2)	0.5938	-25.69	0.91	1.98	0.17	-0.72	0.18	2.12
2	(-2,-2)	0.5990	99.18	69.86	0.39	0.24	6.19	0.26	6.21
3	(2.9,2.9)	0.6437	327.29	45.21	10.47	0.82	41.88	2.05	43.22
4	(4,-2)	1.447	-48.66	16.36	4.16	0.76	-1.94	0.28	4.66
5	(-5,-5)	4.554	-87.01	156.97	-9.02	1.57	-4.55	1.46	10.33

Testpoint 0.

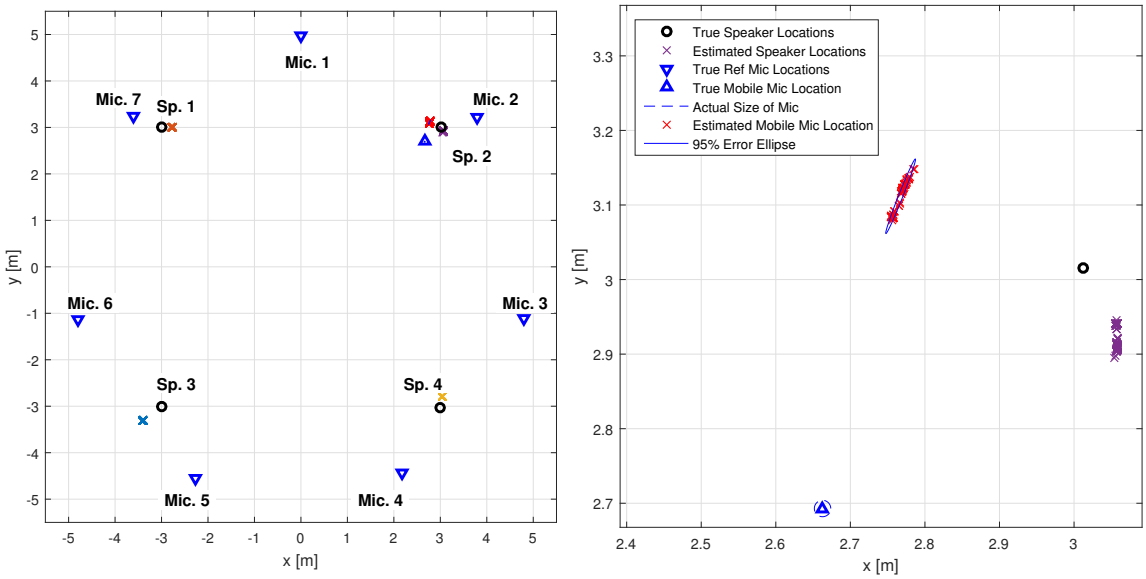


(a) Full test area **(b) Magnified at Testpoint 0**

Figure 47. Estimated location of mobile microphone at Testpoint 0

While the bias present in estimates, shown in Figure 47, is greater than all previous tiers, it is only marginally greater than the bias of Tier 4. The variance, and thus the size of the error ellipse, is consistent with those of previous tiers. The slight increase in bias and consistency of the variance may indicate locating the microphone using simultaneously emitting sound sources slightly degrades the accuracy while maintaining the precision of the estimates. The DRMS of the estimates was 3.00 cm, almost exclusively weighted by the bias.

Testpoint 3.

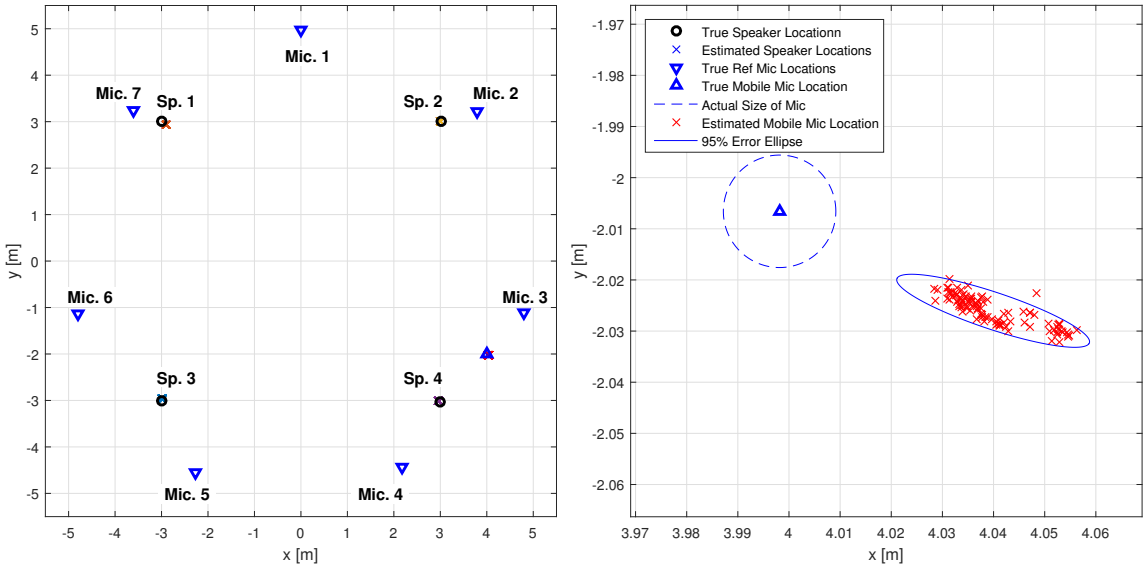


(a) Full test area **(b) Magnified at Testpoint 3**

Figure 48. Estimated location of mobile microphone at Testpoint 3

The test area configuration and results for Testpoint 3 are shown in Figure 48. While results for Tier 4 at Testpoint 3 were significantly biased, the innaccuary of the estimates only grew for Tier 5, producing the highest location estimation errors of all tests. With a DRMS of 43.22 cm, estimates were far above and slightly right of the true microphone location. As shown in Figure 48b, the true mobile microphone location is visually identifiable below Speaker 2, but the estimated mobile microphone location is above Speaker 2. This significant error may be caused by the close proximity of the mobile microphone and Speaker 2. The signals created by the other three speakers could not be separated by the dominant signal from Speaker 2, resulting in errant TDOA measurements. Also of note is the significant error in location estimates for the sound sources, as seen in Figure 48a. The timing estimates were also degraded with a mean clock error of 327.29 μs and a standard deviation of 45.21 μs.

Testpoint 4.

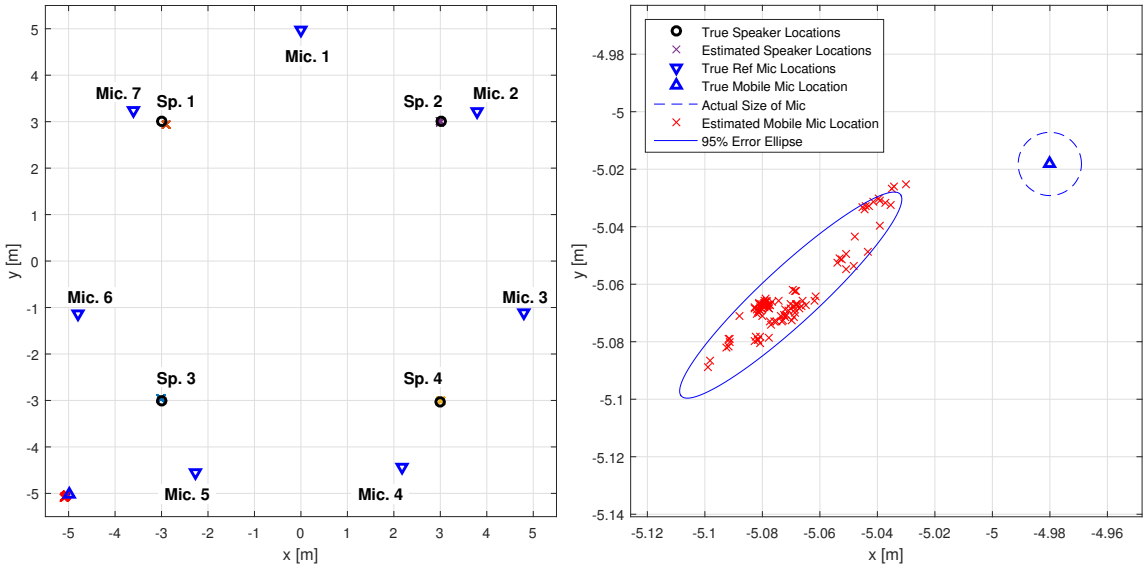


(a) Full test area **(b) Magnified at Testpoint 4**

Figure 49. Estimated location of mobile microphone at Testpoint 4

The test area configuration and results for Testpoint 4 are shown in Figure 49. The DRMS at Testpoint 4 for Tier 5 was 4.66 cm, which was four times greater than the DRMS for Tier 4 with both a greater bias and variance of the estimates as seen in Figure 49b. With the mobile microphone located 1.4 m from Speaker 4, as shown in Figure 49a, it is possible that the estimates were affected by the proximity of the sound source, but to a lesser extent than shown at Testpoint 3.

Testpoint 5.



(a) Full test area **(b) Magnified at Testpoint 5**

Figure 50. Estimated location of mobile microphone at Testpoint 5

The test area configuration and results for Testpoint 5 are shown in Figure 50. Tier 5 produced the highest DRMS value recorded for Testpoint 5 (10.32 cm). While the DRMS is significantly larger than other select testpoints, the error of the estimates is more likely due to the higher DOP of Testpoint 5 than the proximity of a sound source; the closest sound source was 2.8 m away from the mobile microphone. The orientation of the error ellipse is similar to other tests performed at Testpoint 5, but the size was significantly larger.

5.4 Chapter Summary

Tier 5 used TDOA measurements in an LSE to locate the mobile microphone from sound sources of unknown location with unknown signal structure and unknown timing emitting simultaneously. First, likely sound source locations were estimated through SRP. These initial sound source location estimates were used in applying TFM in order to reconstruct the original audio of each sound source. These reconstructions were used in a modified version of the GCC in order to produce the TDOA measurements. Results varied in comparison to previous tiers by testpoint. Testpoint 0 had similar findings compared to Tier 4, whereas Testpoint 3 showed a significant decrease in accuracy of the location estimates. The proximity of Speaker 2 to the mobile microphone caused too much interference when attempting to acquire TDOA measurements from other sound sources.

VI. Conclusion

6.1 Research Summary

The goal of this research was to explore the feasibility of tracking sound as an alternative form of PNT. Specifically, a system was developed capable of locating a mobile microphone using simultaneously emitting SoOPs from sound sources of unknown location. By reconstructing the audio as originally emitted, the system creates clear reference points in the environment. These reference points are used to generate TDOA measurements between the mobile microphone and each of the reference microphones. The TDOA measurements are applied to an LSE algorithm to estimate the location of the mobile microphone. Because of the abundance of SoOPs available in the audible range, the system ideally runs passively, but can also self-generate audio when too few naturally occurring signals are available.

The system was designed in a way to first prove the concept of sound-based positioning, and progressively remove assumptions to approach a more realistic application. The initial design, Tier 1, assumed known timing, position, and waveform of sound sources with sequential playback. Without the need for reference microphones, Tier 1 proved the initial concept of using TOA measurements to estimate the position of the single microphone.

In Tier 2, known signal timing was not assumed, introducing the need for a reference microphone. Comparing the TOAs of the two microphones generated TDOA measurements, which cancel out receiver clock error. Results for Tier 2 showed qualities similar to Tier 1.

In Tier 3, known sound source location was not assumed. With unknown sound source locations, minimum of three reference microphones is required to solve for the location of the mobile microphone. However, to increase the accuracy of location

estimates and to maintain a DOP map shape similar to previous tiers, a total of seven reference microphones were used. The LSE was modified to compensate for unknown sound source positions, which as a byproduct, estimated the locations of the sound sources in addition to the main objective of locating the mobile microphone.

In Tier 4, the system did not assume sound sources generated a impulse waveform, but instead, human speech. The method for determining TDOA values was altered from an amplitude peak detector to the GCC method, which detects time delays in similar signals.

In Tier 5, signals emitted simultaneously instead of sequentially. This difference in assumptions was the most complex threshold crossed toward developing a realistic system. A TFM method was applied in order to reconstruct the audio of the signals as originally emitted and remove the noise created by other sound sources in each channel. The TFM method required known or accurate estimates of the sound source locations. Because the locations were assumed unknown, SRP and PIF were applied to provide an initial estimate of the sound source positions. A modified GCC method, which implemented the reconstructed signals was then applied to provide TDOA measurements. These TDOA measurements were used in a LSE algorithm similar to Tiers 3 and 4 to estimate the mobile microphone position. Results showed evidence of the precise capabilities of positioning via sound. In most cases, estimates were within several centimeters of the true microphone location, if not within the area of the microphone diaphragm. A noticeable contributor to error was the larger DOP values present at some test locations. Error due to DOP can be reduced given a proper configuration of the microphone array.

While not having a foreseeable future of overtaking GPS, the cm-level accuracy provided through passive methods proves sound-based position as a worthwhile area of research in alternative PNT. Positioning via sound also generates byproducts

useful to other applications; producing reconstructed audio of sound sources may be desirable for surveillance purposes.

6.2 Future Research and Applications

6.2.1 Environmental Resiliency.

As an immediate step in progressing the robustness of the system described in this paper, its capabilities could be increased to perform in outdoor environments. Several difficulties arise in outdoor settings. One of which is the introduction of wind, which varies the speed of sound over the test area. A better model of the speed of sound that accounts for wind is required. Another difficulty is the prominence of non-stationary sound sources, which are more difficult to locate using time-averaged SRP maps. For the system to perform well in outdoor environments, the Doppler effect must be considered when locating sound sources with SRP.

6.2.2 Sound Source Detection and Selection Methods.

In estimating the locations of sound sources in Tier 5, it was assumed that there were exactly four sound sources within the test area. In reality, the number of sound sources in a given area is often unknown. In order to apply the solution presented in this thesis to more realistic scenarios, a system should be developed to detect how many distinct sound sources are present and which sound sources would be useful in estimating the location of the mobile microphone. Only three sound sources and three reference microphones are required to form a solution, so additional sound sources and reference microphones not likely to contribute to more accurate results could be ignored.

6.2.3 UAV Detection through Steered Response Power Mapping.

It is also possible to expand the search area of the system upwards to locate airborne sound sources such as small UAVs which may otherwise be difficult to detect. Given the large search volume required for practical application, the computational cost of SRP must be mitigated. Perhaps a trade-off of reduced accuracy to increase computational speed would be permissible. If all reference microphones were located at ground level, the DOP in the vertical direction would be prohibitive to producing accurate location estimates of UAV. Implementing airborne microphones or microphones affixed to towers could reduce vertical DOP. Of course, filtering would be required to ignore sound generated by the microphone host.

6.2.4 Infrasound Positioning for Increased Scalability.

Given adequate performance in outdoor environments, the developed system could be used in large scale surveillance systems. Because of the slow speed of sound compared to the speed of light, sound based positioning is a unique form of alternative PNT. Decreased accuracy in receiver clocks do not have as severe of an impact to sound-based positioning. A 1 ms difference in clocks between receivers translates to approximately a 35 cm error in range difference measurements, whereas a similar clock error for GPS translates to a 300 km pseudorange measurement error. In applications tolerant of meter-level accuracy, several independent receivers, each with its own clock and microphone array, can be combined to create a system capable of covering a much larger area. The tolerance of differences in time errors between receivers would allow use of network clock synchronization though methods such as Network Time Protocol (NTP), while still maintaining fairly accurate position estimation.

Bibliography

1. Yuksel Arslan and Burak Guldogan. Impulsive Sound Detection and Gunshot Recognition. In *Signal Processing and Communications Applications*, 2015.

2. Meysam Basiri, Felix Schill, Pedro U. Lima, and Dario Floreano. Robust Acoustic Source Localization of Emergency Signals from Micro Air Vehicles. In *IEEE International Conference on Intelligent Robots and Systems*, pages 4737–4742, Vilamoura, Algarve, Portugal, 2012.

3. Meysam Basiri, Felix Schill, Pedro U. Lima, and Dario Floreano. Audio-Based Relative Positioning System for Multiple Micro Air Vehicle Systems. *Science and Systems*, 2013.

4. T G H Basten, H E De Bree, and W F Druyvesteyn. Multiple Incoherent Sound Source Localization Using a Single Vector Sensor. *International Congress on Sound and Vibration*, 16(July), 2009.

5. Jacob Benesty. Adaptive Eigenvalue Decomposition Algorithm for Passive Acoustic Source Localization. *The Journal of the Acoustical Society of America*, 107(107):384–391, 2000.

6. J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Location*. MIT Press, Cambridge, MA, 1996.

7. Ramon F. Brcich, D. Robert Iskander, and Abdelhak M. Zoubir. The Stability Test for Symmetric Alpha-Stable Distributions. *IEEE Transactions on Signal Processing*, 53(3):977–986, 2005.

8. H E De Bree. Localization and Tracking of Aircraft with Ground Based 3D Sound Probes. In *33rd European Rotorcraft Forum*, Kazan, Russia, 2007.

9. Hans-Elias de Bree, Jelmer Wind, Erik Druyvesteyn, and Henk te Kulve. Multi Purpose Acoustic Vector Sensors for Battlefield Acoustics, 2011.

10. Aaron Canciani and John Raquet. Absolute Positioning Using the Earth's Magnetic Anomaly Field. *Journal of the Institute of Navigation*, 63(2):111–126, 2016.

11. Andre M. Cavalcante, Rafael C D Paiva, Renato Iida, Alvaro Fialho, Afonso Costa, and Robson D. Vieira. Audio Beacon Providing Location-Aware Content for Low-End Mobile Devices. In *2012 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2012 - Conference Proceedings*, number November, 2012.

12. Tinh Do-Xuan, Vinh Tran-Quang, Tuy Bui-Xuan, and Vinh Vu-Thanh. Smartphone-Based Pedestrian Dead Reckoning as an Indoor Positioning System. In *System Engineering and Technology (ICSET), 2012 International Conference on*, pages 303–308, Bandung, Indonesia, 2012.

13. Kevin D Donohue, Jens Hannemann, and Henry G Dietz. Performance of Phase Transform for Detecting Sound Sources with Microphone Arrays in Reverberant and Noisy Environments. *Signal Processing*, 87:1677–1691, 2007.

14. Kevin D. Donohue, Sayed M. Saghaiannejadesfahani, and Jingjing Yu. Constant False Alarm Rate Sound Source Detection with Distributed Microphones. *Eurasip Journal on Advances in Signal Processing*, 2011, 2011.

15. Joe Khalife, Kimia Shamaei, and Zaher M Kassas. A Software-Defined Receiver Architecture for Cellular CDMA-Based Navigation. In *Position, Location and Navigation Symposium (PLANS)*, 2016.

16. Nicolaj Kirchhof. Optimal Placement of Multiple Sensors for Localization Applications. *International Conference on Indoor Positioning and Indoor Navigation, IPIN*, (October), 2013.

17. Charles H. Knapp and G. Clifford Carter. The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(4):320–327, 1976.

18. J. Krolik, M. Joy, S. Pasupathy, and M Eizenman. A Comparative Study of the LMS Adaptive Filter Versus Generalized Correlation Method for Time Delay Estimation. *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'84.*, 9:652655, 1984.

19. G. II Mellen, M. Pachter, and J. Raquet. Closed-Form Solution for Determining Emitter Location Using Time Difference of Arrival Measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 39(3):1056–1058, 2003.

20. Praveen Reddy Nalavolu. Performance Analysis of SRCP Based Sound Source Detection Algorithms, 2010.

21. Takuma Ohata, Keisuke Nakamura, Takeshi Mizumoto, Tezuka Taiki, and Kazuhiro Nakadai. Improvement in Outdoor Sound Source Detection Using a Quadrotor-Embedded Microphone Array. In *IEEE International Conference on Intelligent Robots and Systems*, pages 1902–1907, Chicago, 2014.

22. Ling Pei, Liang Chen, Robert Guinness, Jingbin Liu, Heidi Kuusniemi, Yuwei Chen, Ruizhi Chen, and Stefan Söderholm. Sound Positioning Using a Small-Scale Linear Microphone Array. In *International Conference on Indoor Positioning and Indoor Navigation*, number 4, Montbeliard France, 2013.

23. Marc Pollefeys and David Nister. Direct Computation of Sound and Microphone Locations From the Time Difference of Arrival Data. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 2445–2448, Las Vegas, 2008.

24. James O. Quarmyne. Inertial Navigation System Aiding Using Vision. In *American Control Conference*, Portland, 2014.

25. Daniel V. Rabinkin. Optimum Microphone Placement for Array Sound Capture. *Advanced Signal Processing: Algorithms, Architectures, and Implementations*, 3162(7):227–229, 1997.

26. Bjorn Schuller, Florian Pokorny, Stefan Ladstatter, Maria Fellner, Franz Graf, and Lucas Paletta. Acoustic Geo-Sensing: Recognising Cyclists' Route, Route direction, and Route Progress from Cell-Phone Audio. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 453–457, 2013.

27. E. Sengpiel. Temperature Dependence of Acoustic Qualities, http://www.sengpielaudio.com Accessed 23 Feb 2017.

28. Spirent. Positioning Application: Can I Get a Dilution of Precision (DOP) Value of Less Than 1?, https://support.spirent.com Accessed 23 Feb 2017.

29. Don J. Torrieri. Statistical Theory of Passive Location Systems. *IEEE Transactions on Aerospace and Electronic Systems*, AES-20(2):183–198, 1984.

30. Harikrishnan Unnikrishnan, Kevin D Donohue, and Jens Hannemann. Time-frequency Masking for Speaker of Interest Extraction in an Immersive Environment. In *IEEE Southeastcon*, Lexington KY, 2014.

31. Xoneca. Geometric Dilution of Precision (Navigation), http://wikipedia.org Accessed 23 Feb 2017.

32. Zhilong Zhang, Weihong Li, and Weiguo Gong. An Improved EEMD Model for Feature Extraction and Classification of Gunshot in Public Places. In *International Conference on Pattern Recognition*, pages 1517–1520, Tsukuba Japan, 2012.

33. Hong Zhao and Hafiz Malik. Audio Recording Location Identification Using Acoustic Environment Signature. *IEEE Transactions on Information Forensics and Security*, 8(11):1746–1759, 2013.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
23-03-2017	Masters Thesis	Sep 2016 – Mar 2017

4. TITLE AND SUBTITLE

Sound Based Positioning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Weathers, David L, 2d Lt USAF

5d. PROJECT NUMBER

16G120

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-MS-17-M-081

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Lab Sensors Directorate, Spectrum Warfare Division,
Navigation and Communication Branch
2241 Avionics Circle Building 620
WPAFB, OH 45433
Email:jacob.campbell3@us.af.mil Phone: (937) 938-4414

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RYWN

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States

14. ABSTRACT

With a growing interest in non-GPS positioning, navigation, and timing (PNT), sound based positioning provides a precise way to locate both sound sources and microphones through audible signals of opportunity (SoOPs). Exploiting SoOPs allows for passive location estimation. But, attributing each signal to a specific source location when signals are simultaneously emitting proves problematic. Using an array of microphones, unique SoOPs are identified and located through steered response beamforming. Sound source signals are then isolated through time-frequency masking to provide clear reference stations by which to estimate the location of a separate microphone through time difference of arrival measurements. Results are shown for real data.

15. SUBJECT TERMS

Sound Based Positioning, Alternative PNT, Time Difference of Arrival, Beamforming, Signal of Opportunity

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. John Raquet, AFIT/ENG
U	U	U	UU	111	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 x4580; john.raquet@afit.edu